

When is a white list another blacklist?

British plans further to tighten the laws on immigration into Britain are either discreditable or are doomed to failure, but either way appear to be designed against the impending general election.

THE British government seems forever to be tinkering with its laws on immigration, chiefly with the intention of restricting the inward flow of people. Part of the explanation is that the social integration of immigrants has been found to be imperfect, to say the least of it. Another is that the financial cost of dealing with immigrants is not negligible. (In the long run, of course, the cost becomes a gain.) But it also seems to matter that a government that is seen to be tinkering with (and tightening) the immigration rules wins the approval of the xenophobic section of the British people, and thus their votes at election time. So it is in accord with recent tradition that the present government, which has to fight a general election within the next eighteen months, should be planning yet another Asylum and Immigration Bill. It is a striking measure of the government's desperation about its electoral prospects that, on this occasion, it has devised a bill that is conspicuously mean-spirited, with elements that belong in the works of Franz Kafka, not in liberal legislation.

The circumstances are these. In considering applications for asylum in Britain, the government now proposes to refer to what is called a 'white list' of countries from which applications for asylum will not be considered. The rationale is that the countries concerned are considered to be 'safe', or free from the threat of political and social repression that justifies the search for asylum elsewhere. That makes sense in certain circumstances. It is, for example, unlikely that people from countries such as Sweden or Switzerland could make a case for seeking refuge in Britain by claiming that their lives would be endangered if they were sent home. But it now seems that the government's white list may include Nigeria, Sri Lanka and Algeria, states hardly distinguished for the humane treatment of their governments' political opponents. (The government confirms the existence of a white list, refuses to specify its contents, but denies the inclusion of the three countries listed.)

The plain truth is that applications for asylum should be judged on their individual merits, and not by reference to arbitrary lists. In 1994, Britain received 33,000 applications for asylum, and adjudicated on 20,990 cases. Britain is far from being an easy touch for asylum seekers: of the 20,990 cases dealt with in 1994, 16,500 were refused, just 825 were accepted and, in the remaining

cases, the applicants were denied refugee status but granted "exceptional leave to remain" in Britain. During the same period, Germany received no fewer than 170,000 applications (and granted a larger proportion of them). In the previous year, Germany took in a substantial proportion of half a million refugees.

The difficulty in Britain is that there is an uncomfortably large backlog of applications, roughly 50,000 in 1994. Those concerned live in reception centres or more generally in the community, now with cut-back social benefits. One of the reasons given for the proposed change of rules is that it will streamline the procedures and so reduce the backlog, but that reasoning carries little force. The mismatch between applications and their adjudication cannot be swollen by people from countries where liberality reigns. Only if the unpublished 'white list' includes countries whose refugees may plausibly claim they will be threatened with persecution if they are returned could its use contribute to the supposed streamlining of the system. But that explanation is denied. The British government cannot have it both ways. Either the white list is as sinister as some suspect, or it is really white in the sense intended, in which case its relevance in the asylum business is unimportant.

Old-fashioned bumbledom is, in any case, a more plausible explanation of the backlog. If the British government were serious about the dangers of immigration, it would long since have done what its European partners have been urging, and would have engaged in the negotiation of common principles to settle the grounds for entry into the European Union (where legitimate residents can move where they wish). In the long run, that is the only way in which to bring this issue to a seemingly conclusion. Sadly, the British government seems to prefer a show of acting tough on immigration for the sake of seeming to act in a manner of which xenophobes would approve. (That is the generous interpretation; the other is that the white list is really a blacklist.) That is the spirit in which the same government rejected a few weeks ago the appeal from Mr Chris Patten, the last colonial governor of Hong Kong, that Britain has a moral duty to allow more than the agreed 50,000 residents of Hong Kong to settle in Britain if they wish. (Most would prefer Canada, with or without Quebec.) The motive can only be to win favour at the impending polls. That is a shameful way to carry on. □



Genomics boosted as Japan unveils plans for state-funded companies

Tokyo. The Japanese government has surprised the country's genetics research community with the announcement that it is to set up two companies to pursue research and development work in genomics. The companies — one established by the powerful Ministry of International Trade and Industry (MITI) and other by the Ministry of Health and Welfare (MHW) — will each be set up in collaboration with private companies from the pharmaceutical, chemical and optical industries. They will be dedicated to decoding genetic information and developing its applications.

Many researchers have welcomed the promise of large-scale government and industry investment in efforts to decode genetic information and develop its applications. But others remain sceptical, questioning whether Japan has enough skilled researchers and technicians to meet the companies' goals, and expressing reservations about the ability of government organizations to navigate a successful path in such a competitive field.

MITI's company, the Helix Institute, will focus on technology for interpreting the function of genetic sequences — so-called second-generation genomic research. The new company will be headed by Teruhisa Noguchi, vice-president of Yamanouchi, one of Japan's largest pharmaceutical companies, and will receive about ¥6 billion (US\$60 million) in investment over six years.

Seventy per cent of this is expected to come from MITI's KEY Technology Center (Key-TECH), a joint venture between MITI and the Ministry of Posts and Telecommunications, set up with interest earned on

money obtained from the privatization of government monopolies, including Nippon Telegraph and Telephone (NTT) and the Japan Tobacco Corporation (JTC).

MHW's company has been provisionally named the Pharma-Genocyte Institute, and will receive about ¥4 billion in investment, also over six years. Half of the money will come from the ministry's Office of Drug-Induced Damages, a fund based on a levy on drug sales. This was originally set up to compensate people who suffered unexpected side-effects from prescribed drugs, but as the government rarely pays compensation, the fund grew so large that changes to the law were enacted six years ago to release the money for venture capital investment.

The president of the Pharma-Genocyte Institute is expected to be Noboru Kobayashi, soon to retire as director of the National Center for Children's Disease in Tokyo. The institute will focus on family studies of the Japanese population to identify relationships between specific genes and disease. According to Yoshiro Ohtaki, of the Japan Associated Finance Corporation (JAFCO), a leading venture capital company and key participant in the new ventures, this may be easier in Japan than in other countries because of well-documented family histories.

The Helix Institute is expected to concentrate on new methods for establishing the function of genes, rather than on genetic sequencing. Ohtaki says that the institute's technical strategy will be decided by a committee formed from representatives of MITI and the participating companies, whose first meeting will be held next month.

Some observers, such as Mitsuru Miyata, chief editor of *Nikkei Biotechnology*, feel that it would have made more sense to create a single company funded by both MITI and the MHW. But Miyata claims that this was made impossible by an "irrational struggle between the two ministries", an example of Japan's traditional inter-ministry rivalry. Communication between the two companies will, however, be facilitated by participants such as Yamanouchi, which will invest in both, says Miyata.

Yoshiyuki Sakaki of the Tokyo University Institute of Medical Science, which has its own genome centre, says that he hopes that the creation of the new companies will help to build up a genome research community in Japan. Indeed, as support for genome research in Japan amounts at present to only ¥2 billion or so a year — concentrated in separate programmes funded by the Ministry of Education Sports Science and Culture (Monbusho) and the Science and Technology Agency (STA) — the new money, amounting to ¥10–11 billion to be invested in the two new ventures over 6 years, will constitute a major shot in the arm for the field.

Sakaki also warns that basic research must not be neglected in the rush to capitalize on practical applications of genome research. But his own institute could directly benefit from MITI's investment, as it is tipped by some as the future home of the Helix Institute. Other possibilities include the Kazusa DNA Institute in Chiba Prefecture, east of Tokyo, and Tsukuba Science City, north-west of the capital.

Those concerned about a possible shortage of skilled researchers include Itaru Watanabe, emeritus professor of Keio University. Watanabe says the lack of qualified personnel can be attributed in part to Japan's university system which, he claims, does not produce enough good postgraduates. He argues that the only long-term remedy would be to create new universities independent of the current system, but suggests that, as a partial solution, the two planned companies should attempt to lure back some of the many Japanese researchers now working overseas.

One academic, while agreeing that 'second-generation' genomic research is an important area to which Japan "must contribute", emphasizes that those responsible for the new programmes must maintain international contacts in order to minimize the duplication of research. "I hope the people who head and organize [these programmes] are aware of the competition in this field," he says. **Stephen Barker**

Earth science museum goes into orbit

London. South Kensington in central London awoke this week to the sounds of construction work on a £12-million scheme to transform the former Geological Museum into what is being billed as "the most exciting and inspiring" Earth sciences complex in the world. A central atrium will introduce visitors to a range of Earth science topics and the forces that have shaped the Earth throughout history, while an escalator will transport them between floors through the centre of an 11-metre-diameter revolving globe (see model, above).

IMAGE
UNAMAGABLE
FOR A CORNABET
FOREASON
REASONS

The project has two phases. A £1-million donation from the RTZ Corporation will help to fund the first phase, due for completion by July 1996. □

Neal Potter Associates

'Science for Peace' offers hope to areas of political conflict

Paris. Daniel Cohen, the French geneticist, last week launched an ambitious initiative designed to promote scientific cooperation in conflict zones. The programme, which would involve building a network of large research centres focusing on biological and genetic research — and would require multi-million dollar funding — is aimed at contributing to political attempts to bring peace and stability to such regions.

To run the project, Cohen has set up a foundation known as "Science for Peace" under the auspices of the United Nations Educational Scientific and Cultural Organization (Unesco).

The foundation has already attracted substantial political support both in France and in countries likely to benefit from the scheme. In particular, the Israeli, Jordanian and Palestinian administrations have formally backed the plans, while negotiations have also begun with Egypt and Morocco. Indeed, political leaders including Shimon Peres of Israel and Yasser Arafat of the Palestine Liberation Organization will take part, with the use of satellite links, in a gala fund-raising concert in Paris later this month supported by Jacques Chirac.

The first of the planned centres would be based in Jordan, where FF3 billion (US\$600 million) would be spent over the next ten years creating a centre for research on drought-resistant crops, with a staff of 1,000 researchers. A second centre, based at Hammamet in Tunisia, would focus on the genetics of malaria and would employ 250 staff at a cost of FF1 billion (see *Nature* 371, 732; 1994).

Cohen plans to create other "technopoles" in Egypt, Israel, Morocco and Palestine, and eventually elsewhere in Africa, Asia and the Americas. But it is not yet clear whether the foundation will be able to attract sufficient resources to accomplish its ambitious plans. Cohen admits, for example, that the foundation has so far only attracted firm commitments for a "derisory" proportion of the FF4 billion needed for the centres in Jordan and Tunisia.

But Cohen argues that, having taken two years to win the required political support, fundraising can now begin in earnest. "Now that this is completed we can begin to look for financing," he says, adding that success of the gala concert will be "crucial". In particular, he is hoping to attract a share of the FF35 billion that will be distributed later this month by the European Commission for the development of the Mediterranean region, and plans to seek further funding from overseas aid programmes, wealthy individuals, and companies.

Declan Butler

Cost-cutting and downsizing take their toll on US R&D

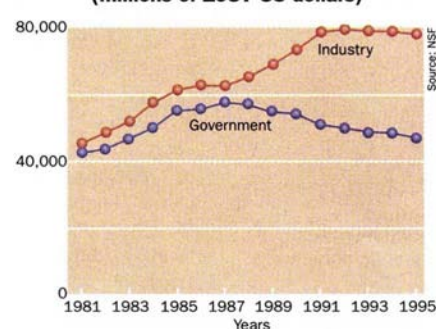
Washington. Total spending by the United States on research and development (R&D) is already in line to fall substantially this year, even before cuts proposed by the Republican Congress get the chance to bite, according to calculations by the National Science Foundation (NSF). In particular, the value of R&D funded by industry will fall by 1 per cent in real terms, and that funded by the federal government by 3 per cent, according to the NSF.

The overall fall means that US spending on R&D as a percentage of gross domestic product (GDP) — the most commonly used indicator of national R&D effort — will fall from 2.51 per cent last year to 2.4 per cent, having slipped steadily from a 1991 peak of 2.8 per cent (see figure, right). Last year, the Clinton administration said that increasing this figure to 3 per cent was "a reasonable long term goal" for science policy (see *Nature* 370, 317; 1994).

The science agency points out some of the main competitors of the United States have also seen a drop in this indicator. In Japan, for example, R&D as a proportion of GNP fell from 2.9 per cent to 2.7 per cent between 1990 and 1993, and in Germany from 2.9 per cent to 2.5 per cent between 1989 and 1993.

These falls, however, have coincided with economic recession, while the United States has experienced considerable economic growth since 1991. "These numbers suggest that we don't stand in a very strong position, even before we put into place the cuts

US spending on R&D by source of funds (millions of 1987 US dollars)



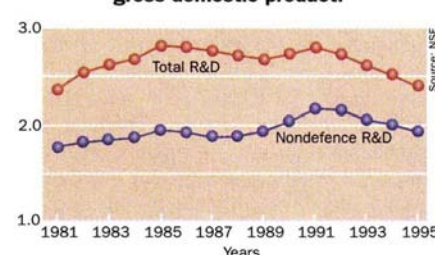
planned by Congress," says Al Teich, science policy director at the American Association for the Advancement of Science (AAAS). Teich attributes the slowdown in industrially supported R&D to the recent emphasis of US corporations on "cutting costs and downsizing".

The NSF estimates that US industry will spend \$101.6 billion this year on R&D, slightly more in current dollars than the \$99.6 billion spent in 1994, but 1 per cent less when inflation is taken into account.

The federal government will spend \$60.7 billion, compared with \$61 billion last year — a relatively small cash drop, but a sizeable reduction of 3 per cent in real terms.

Figures for recent years indicate different trends in support for R&D from government and industry. Government support peaked in 1987, and has since fallen by 20 per cent in real terms (see figure, below)

US spending on R&D as a proportion of gross domestic product.



left). Industry support continued to grow strongly until 1992 — lending credence in some quarters to the idea that it might eventually displace scarce public funding. Since 1992, however, industrial support has also fallen off, despite the strength of the US economy.

According to John Jankowski, programme director of R&D statistics at NSF, the levelling out of research support in industry over the past few years reflects changes in aerospace and parts of the electronics industry, with corporations matching reductions in defence spending with cuts in their own commitment to R&D.

Until recently, basic research spending — most of which is funded by the federal government — has been more resilient than spending on applied research and development. But basic research also fell slightly in both 1994 and 1995, reflecting the fact that, while basic research in universities continues to expand, that financed and carried out in industry has shrunk by 20 per cent from its 1991 peak.

The new figures, which have been released by the NSF's directorate of social sciences, will form the basis for the next major statistical report, Science and Engineering Indicators, 1995, which the NSF's governing body, the National Science Board, will deliver to President Bill Clinton early next year.

The NSF, which has been gathering such statistics since 1953, uses surveys of government departments, universities, companies and other research performers. The Industrial Research Institute helps to collect the information from industry. Expenditures for the current year were extrapolated from information available in June.

Colin Macilwain

US fusion researchers brace for major cuts

Washington. As widely anticipated, the US Congress has made deep cuts in fusion research, renewable energy research and money for technology transfer from the Department of Energy laboratories to industry, in a broad appropriations bill for spending on energy and water which was agreed last week.

The bill — one of the thirteen currently moving through Congress which President Bill Clinton is expected to sign without further argument — sets the budget for the Department of Energy's non-nuclear weapons work for the 1996 financial year, which began on 1 October.

The research programme most severely affected by the bill will be magnetic fusion, whose budget falls from \$375 million last year to \$244 million. This cut marks a historic shift, say fusion researchers, with the United States dropping a 40-year quest for a working fusion power plan to concentrate instead on a slimmed-down effort on basic research into fusion.

The \$244-million fusion budget — \$15 million more than proposed by the House of

Representatives, but \$32 million less than the Senate wanted — will leave the three main US fusion research facilities scrambling for funds, and the future of US participation in the International Thermonuclear Experimental Reactor (ITER) hanging in the balance (see *Nature* 377, 567; 1995).

Anne Davies, the director of the Department of Energy's fusion programme, says she has "no idea" which facilities will continue to run, but promises to decide "within a week or so". She says that the outright closure of one or more facilities will "have to be looked at", and that the "theory and university programmes will be cut too much, and the technology programme just wiped out" by the budget.

Asked about ITER, she says: "I'm sympathetic to ITER but the problems are so bad at all the institutions — we have to figure out a way to do it that the community will not feel to be lop-sided."

The main domestic fusion facilities are the Tokamak Fusion Test Reactor (TFTR) at Princeton, New Jersey, costing \$70 million a year to run; General Atomics' DIII-D

experiment in San Diego, California, which costs about \$50 million; and the Alcator C-MOD machine at the Massachusetts Institute of Technology, which costs \$16 million.

TFTR and DIII-D have not been operating since the beginning of the financial year, pending the budget decision. Tom Simonen, director of DIII-D, says he hopes it will get some money from the "extra" \$15 million to run for eight weeks in the coming year. But, he adds dryly, "there are a lot of demands on that \$15 million".

Also hit hard by the appropriations bill are renewable energy research programmes, which will get \$280 million, down from \$388 million this year. Support for technology transfer from the three weapons laboratories falls from \$225 million to \$160 million and from the non-defence laboratories from \$35 million to \$20 million.

The main physics programmes of the Department of Energy will receive small increases in funding, and \$100 million will be provided for a Special Facilities Initiative to improve the use of physics equipment.

Colin Macilwain

Chernomyrdin promises 'significant increase' for science

Moscow. Viktor Chernomyrdin, the Russian prime minister, promised last week that the government would propose a significant increase in funding for science in its draft budget for next year. At the same time, he called for a major overhaul of Russian science, one of whose objectives would be to eliminate inefficient laboratories "and perhaps even whole institutes".

Speaking at a meeting at the Joint Institute for Nuclear Research (JINR) in Dubna, Chernomyrdin also promised that the government would try to find funds for Russian physicists to participate in the construction of the Large Hadron Collider at the European Laboratory for Particle Physics (CERN).

The prime minister said that the precise figure for next year's science budget is being debated within the government and that it is already clear that the sum will be "significantly higher" than this year's figure.

At the same time, he stressed that there also needs to be more support from industry and the banking sector, and suggested that this should have been encouraged by the recent stabilization of the country's economy. As an example, he announced that the Dubna institute, the Promradtekhbank commercial bank and several Dubna-based military enterprises plan to set up a research and development investment fund to mount joint projects.

Chernomyrdin also said that the legal status and viability of the Dubna institute — one of Russia's leading research institutions

— needs to be preserved, particularly in the light of its recent difficulties (see *Nature* 367, 208; 1994) and promised to provide the funds to allow this to happen.

According to Viktor Mikhaylov, Russia's atomic energy minister, a new agreement with the institute commits the government to regular payment of its annual contribution of about \$17 million a year, which pays for just over half of the institute's running costs.

Mikhaylov acknowledged that delayed payments of contributions from some of the institute's 18 founder states — which include East European and former Soviet countries, Cuba, North Korea, Mongolia and China — had meant that some important research programmes had been halted.

Chernomyrdin's promise of extra funding came shortly after a meeting in Moscow organized jointly by the Organisation for Economic Cooperation and Development (OECD), the lower chamber (Duma) of the Russian Parliament and the Ministry for Science and Technological Policy, the third such meeting over the past two years.

Andrei Fonotov, co-chairman of the seminar and deputy minister of science, told the

meeting that, with elections due in two months time, probably leading to a change of government, it was important that decisions are taken soon on issues facing parliament, enabling achievements to be passed on to whoever takes office in the future.

"We are preparing the drafts of all laws related to scientific activity, constantly trying to coordinate them with ministries and state committees, which sometimes block our efforts for months," said Fonotov. In some cases, even after the government had agreed on a joint policy, parliament had not found time to discuss it; and one of the few laws recently adopted by the parliament — on genetic engineering — has not yet been signed by President Boris Yeltsin on the ground that it is "too declarative".

Mikhail Glubokovsky, deputy chairman of the Duma Committee on Culture, Education and Science — and as such responsible for the science budget — told the meeting that funding for science should be increased at least threefold. "But it will never be done by this government, this parliament and this president," he said.

Glubokovsky produced figures to show that the funding of science in Russia has fallen from 1.03 per cent of GNP (or 3.87 per cent of state budget expenditure) in 1991 to 0.44 and 2.19 per cent respectively last year. The 1995 figure proposed by the government would reduce this still further to 0.34 per cent of GNP, while the Duma had approved a slightly higher target of 0.38 per cent.

Carl Levitin

IMAGE
UNAVAILABLE
IMAGE
UNAVAILABLE
COPYRIGHT
REASONS
COPYRIGHT
REASONS

AP/Alexander Zemlianichenko

Chernomyrdin: pledged higher science budget.

France faces large bill over its space station agreement

Paris. France's Centre National d'Etudes Spatiales (CNES), the world's third-largest space agency, is likely to pay a heavy price for last month's agreement by Europe's space ministers on the participation of the European Space Agency (ESA) in the international space station.

At the meeting, France increased its funding for the European station programme, led by Germany, to ECU756 million (US\$1 billion) over the period 1996–2004 (see *Nature* 377, 669; 1995). In return, Germany agreed to increase its funding for the French-led Ariane launcher programme from ECU120 to ECU292 million.

But to pay for the costs of the station, CNES will need to make economies of 5 per cent annually in its budget — which came to a total of FF10.7 billion (US\$2.17 million) last year — over the period 1996–2000. This could threaten France's position as the only European country able to sustain vigorous national and bilateral space programmes alongside its commitments to ESA.

Anticipating such cuts, CNES staff went on strike on the first day of the meeting, which was held in Toulouse — ironically France's aerospace 'capital' — and held a protest outside the building in which it was taking place. The main concern of unions representing staff at the agency is that the cuts will inevitably translate into job losses both at CNES and within the aerospace industry. Moreover, although France has increased its participation in ESA's contribution to the station, the lion's share of industrial contracts from this programme will go to Germany, which is building the Columbus Orbital Facility (COF).

Investment in France's other space programmes is expected to stagnate as a result of the deals struck at Toulouse, says a representative of staff at Aerospatiale, which last week announced 3,000 redundancies, mainly as a result of cuts in defence spending. The space station, concluded the CNES unions, would have few economic benefits for France, while its costs would weaken the country's space sector and therefore compromise its ability to address future technological challenges.

A similar warning had already been given by the French Academy of Sciences. In a statement issued before the meeting, the academy said that the station was not based on scientific needs; it accepted that Europe might choose to participate for essentially political reasons, but demanded that scientists should not be obliged to use the station and argued that the space science programmes should be protected.

François Fillon, the French minister with responsibility for space, has admitted that

the station's scientific goals are questionable, but that the project's main *raison d'être* is political — namely that Europe could not absent itself from a global space project involving the United States and Russia. But he argues that had the meeting failed to agree on a common European role in the station, Germany would undoubtedly have participated alone, with the other international partners. That would have risked the disintegration of ESA, he asserted, and with it a risk to France's biggest interest in space, the Ariane-5 programme.

Fillon argues that CNES should be able to absorb the cuts with only minimal damage to its programmes. Only one-third of the cuts will be made in national programmes,



he says, whereas two-thirds will be made in ESA programmes other than the space station and Ariane.

André Lebeau, the president of CNES, says that some national programmes will be delayed to make economies, but that the Spot-5 Earth observation satellite and Stentor telecommunications satellite programmes will be maintained intact. French officials also argue that, while CNES will walk a financial tightrope between 1996 and 1998, the situation will then be eased by the completion of the most expensive phases of Ariane-5 development.

But Fillon admits that the effectiveness of any damage limitation exercise will depend on CNES not being subject to further budget cuts. These cannot be ruled out, as the budget committee of the National Assembly has demanded FF240-million cuts in CNES's budget next year.

Whatever the outcome, France's decision to increase its contribution to the space station seems to have put paid to recent efforts by CNES to stop the escalation of its contributions to ESA in order to reinforce its national and bilateral programmes, such as the widely-acclaimed French-US altimetric satellite Topex-Poseidon. **Declan Butler**

Safety protocol due for a rough ride in biodiversity talks

London. Signatories to the United Nations Convention on Biological Diversity will meet in the Indonesian capital, Jakarta, next week amid continued controversy over plans to establish a legally binding international protocol covering the handling, transport and use of genetically modified living organisms.

Representatives of 168 signatories to the UN's Biodiversity Convention will renew calls for a biosafety protocol at their second annual meeting from 6 November. But an advance meeting of biosafety experts specially convened in July by the United Nations Environment Programme (UNEP) failed to arrive at a consensus on the timing of new biosafety legislation.

Representatives of the biotechnology industry have said that they, too, favour loose guidelines on biosafety, or bilateral agreements between countries, such as that signed between the United Kingdom and Argentina (see *Nature* 377, 94; 1995), rather than the formal protocol suggested in article 19 of the convention.

Gavin Cree, for example, group safety adviser at the UK chemical company Amersham International, said last month at a London conference convened by the Royal Geographical Society he could "see no reason" to justify the introduction of a new international biosafety protocol. "Most of biotechnology comes under the same safety regulations as other technologies used in industry and everyday life," said Cree, who was speaking on behalf of the BioIndustry Association.

"The protocol will take time to agree, time to ratify and will not be accepted by all countries. Guidelines can be in place much quicker and will encourage biosafety in countries which lack such regulation."

The Jakarta meeting will also attempt to resolve disagreements about the convention's finances and a proposed 'clearing-house' mechanism for scientific cooperation, as well as the allocation of intellectual property rights. Biodiversity prospecting, where companies pay a tax for the privilege of hunting for an organism as well as a royalty from commercial sales, is beginning to take root in some countries. But this practice does not resolve the issue of ownership of the modified organism.

Tony Juniper, a senior campaigner with the environmental group Friends of the Earth, warns the convention's signatories not to lose sight of the bigger picture as they discuss legal minutiae. "We are losing 76 species every day. Rainforests are under threat. Such emergencies, the real reason for a convention, are drifting further away from importance," he says. **Ehsan Masood**

Industry slow to invest in German biotech...

Munich. Despite considerable political pressure, German industry remains cautious about putting its money into the government's newly launched human genome programme. While supporting the programme, most companies prefer to make their major investments in the United States, drawn by both the more advanced technical base, and by what some see as a more cooperative attitude in the research community.

Negotiations about industry's involvement in the programme have been dragging on about the type of body that should be set up to exploit results from two new academic 'resource centres' — each designed to draw together genetic information and make it available to industry — and on the conditions for such exploitation.

Ministry officials say they are optimistic the negotiations will be concluded successfully by Christmas. But at present, the programme is unlikely to achieve the 50:50 funding split originally proposed by the research minister, Jürgen Rüttgers, when the programme was initially proposed. At its launch last June, he said the programme was only likely to be effective "if industry makes a major contribution".

The genome programme is being set up by the ministry of research (BMBF), and has been promised a total of DM400 million (US\$286 million) over the next eight years. It will include two resource centres, one at the Max Planck Institute (MPI) for Molecular Genetics in Berlin and the other at the German Cancer Research Centre (DKFZ) in Heidelberg.

The MPI has already been awarded DM20 million over two years for work on automated cloning technology based on a hybridization filter technique developed by its director, Hans Lehrach. The DKFZ will receive DM3 million to expand its gene library, verify the data it contains and establish courses in bioinformatics.

The remaining federal funds will pay for competitive grants for individual projects in any area relating to human genome research. Industry is expected to contribute to some of these projects. It will also assist in the setting up of a channel for applying the genetic information identified in the resource centres to the development of drugs and diagnostic techniques.

The programme is already proving popular with the academic community. According to Frank Laplace, the BMBF official responsible for the genome programme, more than 250 grant applications were received by the August deadline. These have revealed in particular a strong interest in analysing the function of gene sequences — rather than in large-scale sequencing itself — and in bioinformatics.

The volume of applications, as well as a complicated system of evaluation which

includes assessment by an international panel, means that the money is unlikely to be distributed before the middle of next year, even though it had been hoped that the programme would start operating sooner.

The delay is expected to allow time for agreement to be reached on industry's relationship with the work of the two resource centres. Originally, the research ministry had proposed that a private company should be set up, fully owned and financed by industry, to negotiate agreements with the centres on behalf of individual companies.

But this model was dropped after running into legal difficulties: in particular, because the service resource centres are supported by public funds, they cannot negotiate directly with outside bodies, but need the agreement of both their federal and regional funding bodies. Furthermore, the companies originally in favour of the model could not agree on how it should operate.

The ministry is now considering a second model, a so-called *Gesellschaft bürgerlichen Rechts*, which would have similar legal status to a private corporation, but whose financing and administration would be shared by the centres and interested companies.

Despite delays in reaching overall agreement, informal consensus has already emerged on some specific points. One, for example, is that any patents arising from research at the resource centres would remain the property of the centres. Another point on which most industrial participants are keen is that they should have privileged access to research results emerging from the two resource centres for three months before they are published.

The companies argue that they need reasonable time to assess the potential value of the research results. But their demand is viewed with concern by the MPI and DKFZ, which feel one month of prior access should be sufficient for a company to establish an interest in a particular genetic sequence.

In addition, not all companies are happy with the way the genome programme limits the choice of sequencing methodology to Lehrach's hybridization filter technique. Many companies would also like to see investment in other types of technology, such as the large-scale sequencing techniques used in the United States, of which several have direct experience as a result of collaborative projects with US companies.

In contrast to their apparent reluctance to invest funds in domestic projects, German companies continue to make substantial investments in gene technology in the United States. Earlier this year, for example, Boehringer Ingelheim announced collaboration worth nearly US\$70 million with Sequana Therapeutics, based in La Jolla, California, which has identified candidate genes for asthma.

Bayer — which is holding back from any investment in the domestic programme — is investing the same sum in Myriad Genetics, in Salt Lake City, Utah, for work on genes related to obesity, osteoporosis and asthma. And earlier this month, Hoechst announced collaboration with Cell Genesys, in California (see opposite), worth up to US\$160 million in the US company's gene therapy programme for HIV infection, based on T-cells, which is now entering phase 2 clinical studies.

Such deals suggest that German industry considers its domestic resource centres to be lagging far behind their US counterparts, and does not regard them as their first choice of research partners. Nevertheless, around 15 companies have already expressed various degrees of interest in participating in the programme. Laplace's main regret is that those least convinced by the programme's priorities are the small gene-based companies such as SoftGene, a bioinformatics company based in Berlin.

Bernhard Wittig, managing director of SoftGene, claims that the genomics programme should adopt a more imaginative approach, including such projects as using artificial intelligence for pattern recognition of DNA sequences, in which his company might be interested.

Given the broader picture, Rüttger's claim that the German human genome programme will "make Germany the leading force [in genomics] in Europe by 2000" appears ambitious. But Laplace remains optimistic that the new programme will provide a solid basis for a domestic genomics industry to grow in the near future.

Robert Unterhuber

Optimism returns

London. After a sustained period of retrenchment, rationalization and falling stock prices, the biotechnology market appears to have recovered its confidence, and may be heading for a new period of sustained growth, according to investment analysts on both sides of the Atlantic.

Last year, the biotechnology industry experienced a number of reversals — including in particular the failure of several promising drugs during the early stage of clinical trials — that led to a collapse in the value of many stocks, particularly those in some smaller companies, and an overall loss in investor confidence.

But this difficult period, which saw a number of large pharmaceutical companies stepping in to provide the experience lacking in start-up operations, seems to have been reversed in the past six months, according to Jeremy Curnock Cook, director of Rothschild Asset Management

...but seeks 'practical' gene therapy deals

San Francisco. Despite a spate of discouraging clinical trials for experimental gene therapies, the interest shown by major pharmaceutical companies in the potential of gene-based medicine does not appear to be waning, according to biotechnology industry experts.

Just two weeks after separate groups of researchers reported that they had so far failed to treat cystic fibrosis and muscular dystrophy successfully with transplanted genes, for example, Hoechst Marion Roussel (HMR) announced a gene therapy agreement with Cell Genesys Inc. worth up to \$150 million, as well as two smaller deals, with the genomics companies Lynx and Incyte, worth a further \$50 million.

These were followed last week by various further announcements, including a \$25-million agreement between Rhône Poulenc Rorer (RPR), which is building up a network of agreements with various research groups in both the public and private sector, and AASTROM Biosciences Inc. for work on gene therapy techniques using lymphoid blood cells.

Executives in these companies remain optimistic about the potential of such alliances, and suggest that the gene therapy reports are unlikely to dampen their enthusiasm. Bob Pearson of RPR, for example, pointed out last week that "the drug industry is used to having failure for many years before eventually success, and it is going to be the same with biotechnology".

In an earlier statement, Jean-Pierre Godard, the global head of HMR, said that the deal in gene therapy for AIDS would place the company "at the forefront of the effort to commercialize emerging and novel

gene therapies".

According to analysts, however, the driving factor behind such investments is not the short-term prospects of rich profits from gene therapy, but the desire to build links with research groups that are able to produce practical clinical data about how disease genes operate. Whether or not they can produce gene-activated therapeutic techniques, small gene therapy companies are seen as a way of providing an efficient path to *in vivo* understanding of such processes.

"The [initial] clinical trials may be negative, but that's not scaring away prospective players, as that's not what they were looking for," said Mark Edwards, managing partner of Recombinant Capital, a consulting company in San Francisco.

Indeed, several major pharmaceutical companies have returned several times to their original partners to set up further joint projects. Baxter, for example, now has three separate agreements with Somatix Therapy, and Sandoz has three alliances with Genetic Therapy Inc.

At the same time, however, the dampening of initial euphoria has helped to introduce a new realism into these arrangements. Thus when Chiron took over the gene therapy company Viagene last month, it announced plans to lay off 20 per cent of the 160 positions at the San Diego company.

Chiron Viagene, the new subsidiary, is in the process of sharpening its focus on a narrower range of potential products, according to its president, Steven Mento. Viagene executives are being encouraged to look upon gene therapy as merely part of a toolbox that also includes small molecules and recombinant drugs. "Gene therapy doesn't

IMAGE
UNAMIGABLE
FOR A COPYRIGHT
FOR A COPYRIGHT
REASONS

New realism? Pharmaceutical companies are continuing to invest in gene-based research.

have to be the magic bullet," says Mento.

Part of the reason for general optimism among pharmaceutical companies is the recent loosening of regulatory requirements for gene therapy trials, and good news of other products in related areas. Linda Miller, an independent financial consultant in the life sciences based in Boston, suggests that their investments will now merely help to accelerate the development of gene therapy, but also benefit from a thinning of ranks of potential competitors.

Some major companies are still holding back from gene therapy as such. Executives with SmithKline Beecham, for example, say that, despite taking a significant stake in genomics through its \$125 million deal with Human Genome Sciences, and although such data are essential for the design of new drugs and diagnostic techniques, the applications to gene therapy remain uncertain.

Others, such as RPR, are continuing to develop strategies that rely on putting together a mix of experts to focus on specific diseases or gene delivery systems. "Companies are beginning to realize that gene therapy breaks down into a number of different elements, including tissue specificity, the targeting of delivery systems, and the regulation of gene expression, and that no one is going to lock into all of the science," says one leading investor.

At the same time, many researchers, perhaps sobered by the recent clinical setbacks in gene therapy, say that they are ready to redouble their efforts to understand the basic mechanisms that will underlie such techniques in future. "The areas of difficulty are clearly identified," says Inder Verma, of the Salk Institute in La Jolla, California. "Now it's just breaking the barriers."

Indeed, if the results of the clinical trials has dampened public hype about gene therapy, this could be a good thing, says Verma. "The public should know the realities," he says. "Not everything is hunky-dory in this field."

Sally Lehrman

to biotech market as stocks recover

in London, suggesting that biotechnology has returned to favour in the market.

Addressing a meeting in London last week of investors in Biotechnology Investments Limits (BIL), a body set up by the late Lord Rothschild in 1981, Curnock Cook said that this optimism was reflected in a 27 per cent increase in the net asset value of BIL's investments — from US\$184 million to \$234 million — between the beginning of June and the end of August.

His analysis is confirmed by figures from investment advisers Piper Jaffray in Minneapolis, which show a sustained growth in the composite value of biotechnology stocks since the early summer, amounting to what the analysts describe as a 'positive trend reversal'.

Other data produced at last week's meeting showed that the recovery in biotechnology stocks has been led by some of the larger and better-known companies,

with 'top tier' companies such as Amgen, Chiron and Genesys all seeing their stock prices increase by more than 50 per cent in recent months.

According to Curnock Cook, one of the main reasons for the recovery has been a growing realization among small companies that the scientific skills required to get a new product into development are not necessarily the most appropriate for handling later stages in the development process, notably the broad-based clinical trials needed for regulatory approval.

Another factor said to have contributed to the recovery in biotechnology stock prices is the desire of investors to find new places in which to put money previously held in high-technology stocks. These have been among the fastest growing sectors of the stock market during the first six months of the year but may now, according to some analysts, be running out of steam. □

Research funds new friends in Japan's Diet...

Tokyo. Japanese science has received a substantial boost in the second supplementary budget to be announced this year — thanks largely to the efforts of a new generation of politicians who are keen to strengthen Japan's science and technology.

The main goal of the supplementary budget is to stimulate the sluggish Japanese economy. But, like the first such budget, announced in May and primarily designed to cope with the Kobe earthquake disaster (see *Nature* 375, 169 & 376, 110; 1995), it also contains hundreds of billions of yen — billions of dollars — for science.

Both moves followed lobbying by some comparatively young members of the Liberal Democratic Party (LDP), the strongest party within the ruling coalition. The key politician pushing for these large budgets for science is Koichi Kato, secretary general of the LDP. Kato is second only in power to the party leader Ryutaro Hashimoto, and is said by one official of the Ministry of International Trade and Industry (MITI) to be "very, very keen" on science and technology.

Another supporter of increased science spending is Koji Omi also of the LDP. In the past, senior LDP politicians have tended to spend their energies lobbying for large construction projects and paying little attention to science. But a series of scandals has trans-

formed Japan's political scene, leading not only to the present coalition government, but also to internal changes in the parties that have allowed younger politicians such as Kato and Hashimoto to come to the fore.

In the latest budget, approved by the Diet in mid-October, the Ministry of Education, Science, Sports and Culture (Monbusho) received an allocation of ¥224 billion yen (US\$2.24 billion) for the promotion of so-

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Universal Pictorial Press, London

Kato: pushed for larger science budgets.

Monbusho's regular system of grants will receive only a small boost of ¥3 billion. But large-scale accelerator projects at the National Laboratory for High Energy Physics in Tsukuba, as well as a large-scale 'helical' fusion device at the National Institute for Fusion Science in Nagoya, will

receive a total of ¥33 billion.

Selected laboratories at some of the larger and better-known national universities, designated as 'venture business' laboratories, will receive an extra ¥11.1 billion in addition to the ¥14.6 billion allocated in the earlier supplementary budget.

MITI will receive an R&D budget of ¥35.6 billion. Of this, ¥18.3 billion will go to the New Energy Development Organization (NEDO), more than half to be spent on a single project aimed at developing ultra-advanced processing technology for the microelectronics industry in which laboratories from both universities and national institutes are expected to participate.

NEDO will use another ¥5.1 billion to continue a grant scheme, inaugurated earlier this year with ¥10 billion from the first supplementary budget (see *Nature* 377, 378; 1995), which is open to researchers at both universities and national research institutes. Other items in the MITI budget include ¥2.6 billion for recycling and other environmental technologies and ¥4.6 billion for a measurements and standards laboratory.

The Science and Technology Agency (STA) has received ¥74.7 billion, including ¥14.9 billion to accelerate completion of its new synchrotron SPring-8 (see *Nature* 377, 566; 1995) and ¥5.1 billion for competitive grants open to all public sector researchers.

Akisha Arima, former president of Tokyo University and current president of STA's Institute of Physical and Chemical Research (RIKEN), says that the new political climate, in which active politicians are pushing hard for increases in the research budget, make him "very happy" — even though he is concerned that "it may not last". Arima has been campaigning for years for dramatic reform of Japan's university research system, which the large new budgets are helping to achieve. **Stephen Barker**

...but CERN fails to gain extra money

Tokyo. Japan's high-energy physicists and the Ministry of Education, Science, Sports and Culture (Monbusho) have failed to pull off a double feat in their attempt to obtain further substantial funding for the Large Hadron Collider (LHC) at the European Laboratory for Particle Physics (CERN) in Geneva, Switzerland, in the latest supplementary budget (see above).

Earlier this year, Japan gave CERN a pleasant surprise when it announced plans to contribute ¥5 billion (US\$50 million) towards construction of the LHC (see *Nature* 375, 169; 1995). Scientists and the ministry had both been hoping to supplement this with a further contribution of ¥2 billion towards a "common fund" being set up for the ATLAS detector on the LHC. CERN has asked both member and non-member participants to make contributions to the common fund as soon as possible, in addition to their own individual budgets for participation in ATLAS.

Monbusho held preliminary discussions with the Finance Ministry about including the ¥2 billion in last month's supplementary budget. But, say Monbusho officials, it never got as far as an "official request".

Keisuke Yoshio, Monbusho's coordinator for international programmes in the international science division, says the ministry can

still apply for the ¥2 billion at a later stage. One such possibility would be the budget for the fiscal year 1997, requests for which will be made next August. But the ministry had been hoping for a quick promise of the ¥2 billion in order to demonstrate its commitment to CERN and ATLAS.

Meanwhile, the education ministry has made a small budget request of ¥92 million for fiscal 1996 (which begins in April) as a first step towards obtaining a total contribution of several billion yen for ATLAS between now and 2003. A decision on that request will be made by the Finance Ministry and the cabinet at the end of December. Such requests are seldom rejected, although they can be trimmed back.

Masayuki Inoue, director of Monbusho's international science division, who has led the negotiations with CERN, admits that even the ministry was slightly surprised when it won ¥5 billion in the first supplementary budget in May. That budget was a response to the Kobe earthquake disaster, and to explain the windfall for CERN he quotes an old Japanese saying "*ten no toki, chinori, hito no wa*" — which roughly translates "even when the heavens provide an opportunity, and conditions on the ground are right, human cooperation is the key to success". **David Swinbanks**

AAAS calls for end to 'ex-communists' ban

Washington. A report from the American Association for the Advancement of Science (AAAS) has called on governments in former communist-ruled countries to end blanket discrimination against academics alleged to have had ties to communism.

The report, released in Washington last week, suggests that so-called 'lustration' (cleansing) laws, which have resulted in large-scale dismissals of university staff in Bulgaria, the Czech Republic and Germany, should only be implemented on a case-by-case basis. The AAAS also calls on post-communist governments to implement methods of democratic reform "that do not contain repressive elements". □

HIV-blood scientist defends reporting delay

Washington. A consultant virologist hired by a US manufacturer of HIV-infected blood products has defended himself against criticism for not going public sooner with the results of research carried out in 1985 showing the potential inadequacy of the company's heat treatment for killing the virus.

At least six haemophiliac children became infected with HIV after the company, Armour Pharmaceutical — now a subsidiary of Rhône Poulenc Rorer — one of the top five blood products companies in the US, continued to market its blood clotting product, Factorate, in Canada for two years, despite research showing that not all traces of HIV were destroyed by its heat treatment.

The research had been conducted by Alfred Prince, head of the laboratory of virology and parasitology at the New York Blood Center. Prince argues that Armour's decisions not to withdraw the product, nor to inform the Food and Drug Administration (FDA) immediately of his results, were justified in the light of uncertainty at the time about HIV.

Prince denies that the company was motivated by a desire to maintain sales in the face of stiff competition. "Actions taken then cannot be judged in the light of our knowledge today," he says, adding that, as the process of removing Factor VIII from plasma involves a number of steps which could remove or activate the virus, "we pointed out at the time that the relatively modest virus inactivation resulting from the Armour heating process did not necessarily imply that their actual product was unsafe".

But haemophiliacs in both the United States and Canada claim that company documents made public for the first time last month suggest Armour was also motivated by a desire to avoid the delays and extra investment needed to introduce an extended heat treatment.

"They had virus [in the product] and Prince knew it; that is not a grey area," says Corey Dubin, vice-president of the US-based Committee of 10,000, an activist group of haemophiliacs infected with HIV, and a member of the FDA's Advisory Committee. Dubin claims that by 1985, it was well-known that HIV infection was transmitted by blood, and that contact could lead to AIDS.

Over half of 16,000 haemophiliacs in the United States have become infected with HIV through contaminated blood products, and three hundred have sued Armour. Up to now, the documents surrounding the suits have been sealed by US courts. But the Canadian action means that they have been made public for the first time.

The documents reveal that an active debate took place within the company — then owned by Revlon Health Care —

following Prince's conclusions that only 70 per cent of virus added to blood products was killed by Armour's method of heating its blood products, which involved keeping them at 60°C for 30 hours.

Despite the indication that not all the virus had been killed, Armour officials felt that Prince's results were insufficient to halt distribution of Factorate, given that the FDA had approved the heat treatment method and the product itself the previous year. "We could only say that the process was not strong, not that the process was unsafe," Prince said.

But the report of an internal company meeting indicates that company executives were also reluctant to commit themselves to the extra investment needed to introduce a more costly process of heat-treating the product, used by its major competitors, keeping it at 68°C for 72 hours.

The minutes of Revlon's Recombinant DNA Steering Committee, held on 15 October 1985, reveals several executives arguing against informing the FDA of the preliminary results of studies showing that "detectable levels" of HIV remained after heat treatment. Some stressed the need to remain competitive in the face of "bids coming up in Canada and France".

The minutes of the meeting also appear to confirm Prince's complaint that he wanted to publish his results, but was prevented at the time from doing so by Armour, one of whose executives agreed to "review the relevant agreement with Prince to ensure publication does not violate our proprietary data rights".

Prince subsequently repeated the experi-

ments independently, and published the result in *The Lancet* in 1986 (1: 1280, 1986). In a statement issued last month, Prince says that "as Armour had previously been informed by the Centers for Disease Control and Prevention (CDC) and the FDA that their process was highly effective, our data were understandably greeted with initial scepticism."

Despite the results of Prince's study, Armour proceeded with its plans to sell Factorate to the Canadian government's health system. In January 1987, apparently in response to growing concerns, the company began selling higher heat-treated Factorate in January 1987. But its distribution of the lower heat-treated product was not halted in Canada until November 1987.

Some critics, such as Dubin, say that Prince should have gone directly to the FDA, and been more vocal about his results. "Whether anybody sues Prince personally is still open," says Dubin.

But Mark Feinberg, a virologist and member of a committee at the Institute of Medicine that had just completed a study of the US blood system, says that even if Prince had spoken with the FDA, the agency probably would not have informed other blood products companies and the public of the research results, in order to protect Armour's proprietary interests.

Harold Sox, chair of both the department of medicine at Dartmouth Medical School and the Institute of Medicine panel, suggests the FDA would not have withdrawn the product in the belief that "companies would not knowingly distribute product infected with HIV".

Adrienne Appel

Accidental contamination 'unlikely'

Washington. The US Nuclear Regulatory Commission (NRC) says accidental contamination is unlikely to have been the cause of an incident at the Massachusetts Institute of Technology (MIT) in August in which a postdoctoral researcher ingested a large dose of radioactive phosphorus-32 (see *Nature* 377, 563; 1995).

At a meeting held to close the first stage of the commission's investigation of the incident, John Glenn, an NRC inspector, is said to have told David Lister, MIT's vice-president for research, that no evidence had been found of an accidental spill of the radioactive tracer.

But Glenn said he could not entirely rule out an accident. According to Victor Dricks, a spokesman for the NRC, a team from the commission's Office of Investigation will continue to investigate the circumstances in which the radioac-

tive material was ingested.

The NRC also reported the discovery of "some problems with the security" of radioactive elements at the laboratory of Susumu Tonegawa, where the incident took place. Ken Campbell, a spokesman for MIT, said that the main problem encountered by inspectors was the use of a chair to hold open permanently a door that the law requires to be locked.

The NRC is now preparing its own estimate of the amount of phosphorus-32 ingested by the researcher, Yuqing Li. MIT officials say that the dose was 579 microcuries — just below the 600 microcuries that would have required it to notify the commission when the incident took place. MIT notified the NRC on 16 October, when reports were about to surface in the press. A third estimate is being prepared independently at Li's request.

Colin Macilwain

Deterioration of Down House

SIR — You have given a misleading impression of the involvement of the Royal College of Surgeons (RCS) in Down House, Charles Darwin's home (*Nature* 377, 90; 1995).

You fail to mention a number of important points. Following the death of Darwin's widow in 1896 and until the 1920s, the family permitted the use of Down House as a school. Its future was a constant problem and it was eventually sold in 1929 to Sir George Buckston Browne who entrusted its care as a national memorial to the British Association for the Advancement of Science. The association maintained the property for many years, but found the running costs insupportable and, in 1953, the property was offered to the RCS.

Since then, the college has spent several hundred thousand pounds to maintain the house while striving with limited funds to fulfil its primary responsibilities in the professional fields of research, education and the maintenance of surgical standards. The National Trust was approached many years ago with a view to taking over Down House but declined. Three years ago, the Natural History Museum agreed to take it over on a 99-year lease and set about raising the funds necessary to restore it and up-date its facilities for visitors. It has failed to do so and a very substantial shortfall on its target is now, I understand, being sought from the National Lottery Heritage Fund. I would add that, over the past three years, there has been a visible deterioration in the external condition of the property.

Your allegation that "the rot set in" following the transfer of this property to the RCS is incorrect.

Rodney Sweetnam

(President)

Royal College of Surgeons of England,
35-43 Lincoln's Inn Fields,
London WC2A 3PN, UK

■ Contributions to the Down House Appeal may be addressed c/o Natural History Museum, London SW7 5BD. □

Contamination

SIR — In response to the contamination of a researcher at the Massachusetts Institute of Technology (MIT)'s cancer laboratory (*Nature* 377, 563; 1995), the US Nuclear Regulatory Commission (NRC) has despatched an Incident Investigation Team to the facility. As part of its investigation, NRC will prepare an incident chronology; identify the source and nature of the phosphorus-32; project actual and potential dose consequences; evaluate the licensee's event reporting and response to

the incident; investigate potential wrongdoing at the laboratory; and determine whether the NRC's regulatory procedures and activities preceding the event may have contributed to it.

The NRC has also sent a letter to David J. Litster, MIT's vice-president and dean for research, requiring that no later than 24 October certain steps be taken, ensuring among other things that security over radioisotopes is adequate to provide reasonable assurance against another such incident. Meanwhile, the NRC investigation of the contamination event that occurred in late June at the National Institutes of Health in Bethesda, Maryland, is continuing. The NRC does not, by the way, dispute the pregnant researcher's contention that she received a higher dose of P-32 than 26 other researchers contaminated at the laboratory.

Victor Dricks

US Nuclear Regulatory Commission,
Office of Public Affairs, Region I,
475 Allendale Road,
King of Prussia, Pennsylvania 19401, USA

Good manners

SIR — The commentary by John Maddox on "Restoring good manners in research" (*Nature* 376, 113; 1995) may have been fitting for his address to a gathering of science editors, but it was wide of the mark as regards a description of the real crisis in the research community. While we should not accept increasingly frequent "bad manners" in publishing, I do not agree that academic institutions and grant-making agencies have primary responsibility to ensure against these unfortunate activities. As in other such institutions, faculty at this university submit thousands of manuscripts each year. As the chief institutional research officer, can I ensure that the bibliography of each is accurate, not self-inflated, or appropriately inclusive of others? Clearly not. Institutions and agencies cannot and should not get into the business of censorship.

Most importantly, these matters should not distract us from more fundamental problems — a shocking increase in the number of cases involving significant alterations of data to suit one's preconceptions, misappropriation of scientific information for commercial gain, falsified effort or outright fabrication of data. In recent years, I have witnessed wholesale and egregious dishonesty in research, a different situation from the 'old days', and, I hasten to add, a picture that is not unique to this university. It is research misconduct of this type where academic institutions and agencies must take primary responsibility for establishing limits of

acceptable research behaviour with comprehensive policies as well as mechanisms to ensure penalties for those who violate them.

Richard K. Koehn

210 Park Building,
University of Utah,
Salt Lake City, Utah 84112, USA

HIV and AIDS

SIR — The recent study of Darby and colleagues¹ provides a direct link between infection with HIV and increased mortality due to AIDS in British haemophilia patients. Using a different approach, our group has studied the incidence of AIDS in the Sydney blood transfusion population (all infected before the introduction of screening for HIV in 1985) and found a direct association between the HIV antibody status of the donor and the incidence of AIDS in recipients².

A successful donor/recipient Lookback Programme initiated by the New South Wales Red Cross Blood Transfusion Service in 1986 has identified specific HIV-positive donors for 101 out of a total of 132 HIV-positive blood transfusion recipients. In cases where the donor was identified only because a recipient became HIV-positive or developed AIDS, the first case of AIDS was excluded to remove selection bias. Excluding such cases, 66 blood transfusion recipients were traced of whom 40 have developed AIDS. In 1988 a parallel investigation aimed at finding blood transfusion-related cases was begun, and 7,000 recipients were tested of whom 10 were found to have AIDS. When these two groups are compared, the relative risk of AIDS associated with receiving blood known to be HIV-positive from tracing, is 424 (95% confidence interval 342-526, $P < 0.00001$) (ref. 2).

Although the study of Darby *et al.* provides more impressive data in terms of sheer numbers, our transfusion study is unique in that (1) we have information on the health status of the donors and recipients, (2) we can link specific donors and recipients and (3) dates of infection are documented.

Collectively, our study² and that of Darby *et al.* provide important and persuasive evidence directly linking HIV-antibody positivity with progression to AIDS.

John S. Sullivan

Jennifer C. Learmont

Andrew F. Geczy

Wayne Dyer

NSW Red Cross Blood
Transfusion Service,
153 Clarence Street,
Sydney, NSW 2000, Australia

1. Darby, S. *et al.* *Nature* 377, 79-82 (1995).

2. Sullivan, J. *et al.* *AIDS Res. hum. Retrovir.* 11, 1147-1148 (1995).

Radioastronomy and the unquiet radio sky

This week's meeting of the World Radio Conference could make or break radioastronomy by making wise or foolish decisions about the reservation of radio-frequencies.

Westerbork. The easy way to gauge the seriousness of the damage done to radioastronomy by the general radio traffic in the sky is to stand in front of one of the pen recorders at this radioastronomy observatory in North Holland (a few kilometres south of Gröningen). You can quickly tell when people wake up and move about in this part of the world (soon after 06:00): the pen recorders quickly go off scale. Thus has radioastronomy been converted from a 24-hour-a-day business to one in which, like optical astronomers, observers are also compelled to stay up all night.

Westerbork is the site of the array of radiotelescopes originally built by the late Jan van Snort and still in service after nearly 30 years. The telescope is still widely used (not only by Dutch astronomers), but is busily rebuilding its equipment to make the frequency of its receivers more easily interchangeable. The hutted buildings are stacked with double-frequency receivers that will soon be bolted to each of the 14 dishes, and which will make it possible to switch from one frequency to another in a few hours, not a few days as at present.

By accident, the observatory has also been the cockpit of the fight against contamination of the radioastronomy spectrum waged by the research community against the commercial users of the radio spectrum over the past several years. And the accident is not really accidental. H. C. Kahlmann was the natural choice as chairman of the European Science Foundation's Committee on Radio Astronomy Frequencies. Everybody is hoping that the committee's report, published a few weeks ago, will have some effect on the current negotiations at the World Radio Conference, now under way at Geneva.

'Commercial' is the word that matters. One wag has calculated that if radioastronomy were to sell off its present allocation of protected radio frequencies to commercial interests at something like the prices for which frequencies have recently been auctioned in the United States, there would be enough cash to provide every radioastronomer with an income of \$160,000 a year for the rest of time. And it would be a life of leisure; radioastronomy as a science would have to fold its tent.

There are also rumours, unconfirmed, that companies with interests in radio-telephony are seeking to make deals of unspecified character with particular observatories, or suborning individuals by offering them lush consultancies. These

tales are a sign that radioastronomers appreciate how much they are asking of the real world by demanding that certain frequencies should be reserved for them alone. But they appear also to recognize that they are playing a poker game. In a strictly rational world, they would put the frequencies they wish to see reserved in some kind of order of priority — and would then find them whittled away from the bottom. Better to keep the pirates guessing.

What the Kahlmann report does instead is to give a thumbnail justification in scientific terms of the reasons why particular wavelength bands deserve protection. The tone of the document is far from defensive, which is right and proper. Recent years have enormously extended the number of particular frequencies at which the observation of emission from the sky can be of immense interest to astronomers at large. In the 1950s, there was only the 21 cm (1,420 MHz) line of hydrogen and, later, that of the molecular inversion of ammonia. Now there are not merely CO and a host of other molecular species, but the search for OH masers (at 1,665 MHz), the need for quiet bands in which to search for redshifted versions of the 21-cm emission from neutral hydrogen and the wish to learn a little more about quasars from the freedom to use measurements in the 15.40 GHz band. Will this year's World Radio Conference grant all these wishes, and on what terms?

The first thing to note is the magnitude of the claim that radioastronomy is making of the commercial world. The companies bidding to use the protected frequencies would not willingly contribute to research in astronomy the many hundreds of millions of dollars a year they would pay for the right to use the same frequencies. But that is also why the radioastronomers' demands are reasonable.

There is no method by which some latter-day Solomon could carry out an accurate cost-benefit analysis on behalf of radioastronomy and set that off against the commercial value of a free-for-all in frequency use. Nor would it make sense to let the market decide, presumably by inviting research councils and other grant-making agencies to bid for frequencies in competition with the companies. Even if they were able to get their bids in on time they would invariably bid too low, at least the first time around. Then some kinds of observations would be out of bounds until the next

round of bidding, perhaps decades away.

The only acceptable route to a solution is that governments should openly acknowledge that the needs of radioastronomy, esoteric though they may be, are in some sense paramount. It will be ridiculous if a few years before the end of this century and millennium, one crucial window on the still largely unknown Universe will be closed or at least clouded by foolish decisions hastily taken. When most well-off governments are wringing their hands about the preservation or their cultural heritage, that would be a shabby business.

For what it is worth, there is no permanent protection from radio interference in long baseline interferometry, although the direct effects are rather less severe; to the extent that radio contamination is local or, at worst, regional, the comparison of signals from distant receivers can be used to cancel out the contamination. But in the processing of signals, there is no such thing as a free lunch. The signal-to-noise ratio of a combined signal is always degraded by manipulations meant to remove contamination.

So much will soon be apparent at Dwingeloo, a few miles down the road to Amsterdam from Westerbork. This is the site of the European very-long-baseline-interferometry (VLBI) correlator station, which now exists as a large hole in the ground near the steerable dish built here in the 1960s. The project has been on the boil since 1988, when the European Community (now Union) funded a feasibility study (and then declined to follow through).

Since then, arm-twisting (in which Nature played a part) has unlocked the funds for the building from the same source, while national governments have done their duty by their own people in contributing to the cost of the small core staff. The main telescopes in the club contribute up to 20 per cent of their time on VLBI work. The network stretches from the eastern Pacific to China and Japan, and has 16 regular participants. It may soon be possible to work with magnetic recording tape a few micrometres thick: that will make it possible to cram 24 hours of recording onto a single reel. And there is great excitement at the planned launch next summer by the Japanese Institute of Space and Astronomical Science of the first steerable radiotelescope in orbit, which will become another node in the network.

John Maddox

Perspectives past and present

Philip Newton

UNDERSTANDING the intricate web of processes controlling interactions between the Earth's climate and its ecosystems is a daunting task, but an essential one if we are to play any kind of informed role in minimizing climate and ecosystem perturbations due to anthropogenic activities. Intergovernmental policy decisions on practices that carry a perceived threat to climate or ecosystem stability have been — as is clear from the present prevarication on CO₂ emission policy — largely based on balancing a desire for minimal perturbation against short-term economic costs; we simply do not yet have a sufficient physical, biological and chemical understanding of the integrated 'Earth system' to argue otherwise.

A conference in September*, convened by the International Geosphere-Biosphere Programme (IGBP), aimed to serve as a focus for steering future research towards the development of dynamic biogeochemical ecosystem models that interact realistically with the physical ocean-climate-land system. What emerged was a melting-pot of different models, rapidly evolving as a result of being continually constrained by data, yet liberated by ever more ingenious computational techniques. At this rate, the ability to predict the broad characteristics of regional climate — and the anthropogenic contribution to it — on seasonal to interannual timescales could be realised within the next ten years.

Our entire understanding of the functioning of the Earth system is based on observations. Characteristics of the Earth's past climate have been preserved in the geological, ice and vegetation records, noted by historians and, over the past few decades, measured directly. The varying timescales and resolutions of these observations dictate the timescales on which the Earth system can be modelled. The conference was tailored accordingly, being divided into sessions on the palaeo, historical, recent and future eras.

Changes over the past 200,000 years, encompassing the last ice-age and the preceding interglacial period, can give us a perspective on the functioning of the Earth system under natural forcing, such as changes in the Earth's orbit and ocean circulation. Studies over the past 2,000 years enable the growing influence of mankind to be followed, first as land-use changes (agriculture, deforestation, irrigation), then as industrialization (largely affecting trace gas emissions) over the past 150 years. On a yet shorter timescale,

the past few decades have provided the most observations as well as being a period of relatively rapid change in atmospheric composition and land use.

Participants at the meeting were treated to a fascinating range of perspectives on our present 'best-guess' as to how the Earth system has changed and responded to change over each of these timescales. On the key question of what is happening now, and what will happen in the near future, the evidence is mounting that the elusive 'missing sink' of carbon (the fraction of anthropogenic CO₂ emissions that cannot be accounted for by present estimates of atmospheric, oceanic and terrestrial sinks) lies predominantly in the terrestrial biosphere. The ability to measure gas concentrations and their isotopic signatures with unprecedented accuracy now allows the detection of variations in the O₂/N₂ atmospheric ratio (R. Keeling, Scripps Inst. Oceanography) and the isotopic composition (for carbon and oxygen) of atmospheric CO₂ (P. Ciais, Commissariat à l'Energie Atomique). Measurements at various locations are already under way, and their continuation should allow the various CO₂ sinks to be located and quantified ever more accurately as the time-series data accrue.

Achieving this goal is crucial, as, if we cannot describe the present-day global carbon cycle, there is little hope of predicting what the future holds. And even if we can balance the global carbon budget, we must also understand the underlying processes. How robust is the terrestrial sink for anthropogenic CO₂? Will it saturate, and if so when, and how will it be affected by global change, such as changes in vegetation or nutrient cycling? Indeed, there were intriguing allusions (S. Piper, Scripps) to preliminary data analyses that suggest that the large terrestrial sink of anthropogenic carbon in the Northern Hemisphere may have been much less significant 10 years ago. If so, we may already be seeing a dynamic and evolving ecosystem response to global anthropogenic perturbation.

In general terms, we have a better understanding of the functioning of the atmosphere than of the ocean, and of the ocean than of the terrestrial biosphere. This ranking is reflected by the sophistication of existing models. The terrestrial biogeochemists are trailing because of a comparatively late start and what is probably a more complex system to model. But the first dynamic biogeochemical ecosystem models, aiming to simulate the full range of ecosystem responses to changing climate — such as vegetation changes or migration, and perturbations to the global

biogeochemical and hydrological cycles — are just starting to emerge (F. Woodward, Univ. Sheffield; J. Melillo, Woods Hole Oceanographic Inst.; A. Friend, Inst. Terrestrial Ecology).

So where do we go from here? The half-dozen or so model intercomparison exercises presented revealed that, in general, global analyses are in better agreement than regional analyses, but that the discrepancies are large. The different models need to be better constrained by data in order to produce more accurate outputs. But global data sets are expensive, and resources are few, so we must invest wisely.

What are the right kind of data? One example of how to go about answering this question was given by I. Fung (Univ. Victoria). We are most likely to find direct field evidence for (and thus be able to quantify) the 'missing sink' in the ecosystems where the sink is both detectable and statistically significant compared to that due to natural variability. She applied a simple carbon model of the biosphere forced by 100-year data sets of temperature, precipitation, nitrogen deposition and atmospheric CO₂ concentrations to show how one might, in principle, optimize site selection and sampling strategy. Her analysis suggested that we are most likely to be able to make unequivocal observations of the 'missing' terrestrial carbon sink if we look in the soil carbon, over timescales of up to ten years or so, and where — geographically — to look (suitable sites seem to be few in North America and western Europe). This particular sampling strategy should not be taken too seriously, as the model lacks sophistication, but the principles of the approach were met with enthusiastic approval.

In the short term, there is the goodwill to share available data archives, to undertake more rigorous model intercomparisons as part of the process of model evolution, and to anticipate collective data needs. Can we achieve the goal of reliably predicting the broad characteristics of regional climate on seasonal to interannual timescales within the next ten years? G. Asrar (NASA) firmly believes so, and that the required remote-sensing framework is already in place. Predictions on decade timescales would be a few years further down the line, and those scientists working on what is possibly the critical piece in the jigsaw — modelling a dynamic interactive terrestrial biosphere — are not alarmed by this timetable. It seems that the expertise is in place and willing, as is the guiding framework of the IGBP. The key to success will probably be making the right measurements at the right places at the right times — and, as always in science, having the resources to do so. □

Philip Newton is an assistant editor of *Nature*.

*Global Analysis, Interpretation and Modelling: First Science Conference, Garmisch-Partenkirchen, Germany, 25–29 September 1995.

Maps of time and space

Mriganka Sur

CRUCIAL to our perception of the world are orderly maps of the body's sensory surfaces in the brain. For most sensory pathways, such as those involved with sight, touch or hearing, an important principle is the maintenance of this topography, whereby adjacent sites on the receptor surface project to adjacent sites in brain centres. How does the brain create and maintain these maps, and what do they signify?

On page 71 of this issue¹, Merzenich and co-workers provide compelling evidence that maps in somatosensory cortex are spatial memory-traces of the timing of tactile stimuli. The cortex uses the temporal coincidence of arriving inputs to maintain and even create spatial representations—inputs that are synchronously active are represented together and integrated right down to the level of single neurons, while asynchronous inputs are actively segregated.

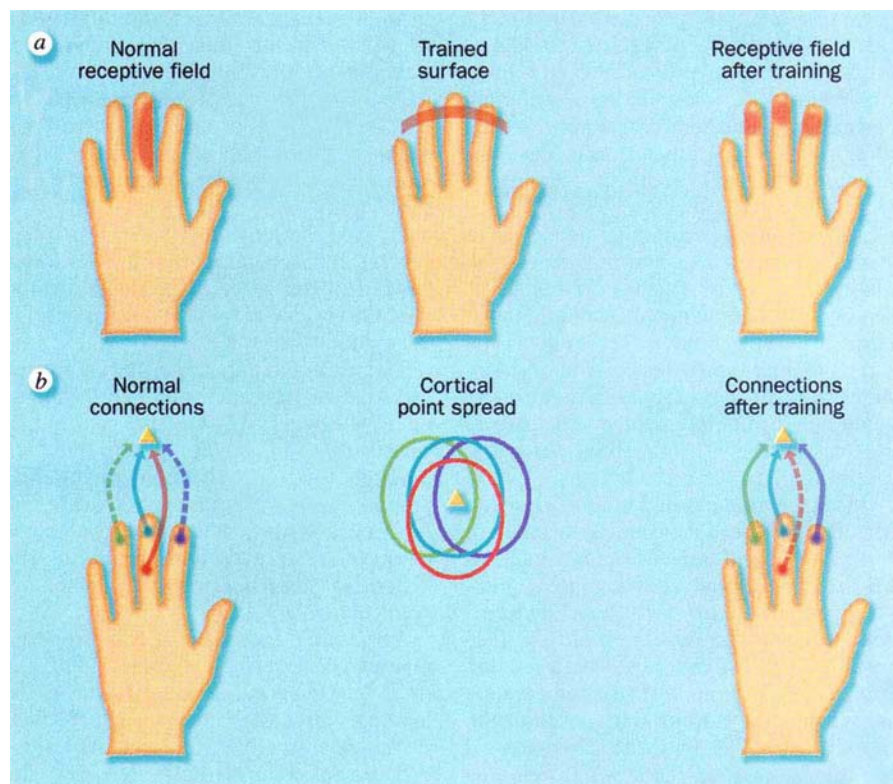
It has been known for some time that space is not the only parameter that comes to be mapped in orderly fashion in the cortex. In the primary visual cortex, for example, neurons respond (depending on the species) to various stimulus attributes such as line orientation, direction of movement, eye dominance, spatial frequency, colour and stimulus contrast. There are systematic and partially continuous maps of these attributes horizontally across the visual cortex, grouping together neurons that prefer a particular value of one attribute and representing them next to neurons that prefer an adjacent value². Although these representations are created during development, a great deal of evidence (derived mainly from studying maps of the body in somatosensory cortex) indicates that they are maintained dynamically and can be altered by changes in inputs³.

The experiment reported by Wang and colleagues¹ is simple in principle and elegant in practice. Normally, the representation of the hand in somatosensory cortex of monkeys contains neurons with receptive fields confined to a single finger (shown in part *a* of the figure). Wang *et al.* reasoned that such receptive fields, and the accompanying discrete representation of each finger in cortex, occur because normal use of the hand by monkeys (and humans) causes synchronous stimulation of skin mainly on one finger and asynchronous stimulation of adjacent fingers. They trained adult owl monkeys on a task that involved stimulation in a direction orthogonal to normal use, using a bar that stimulated either the distal or proximal segments of three fingers simultaneously. They found that

many neurons now had receptive fields that spanned the stimulated surfaces, including several fingers. Furthermore, there was a discontinuity in the map between the distal and proximal digit representations, rather than between individual fingers. In other words, the spatial properties of neuronal receptive fields and the spatial features of the map came

The answer lies in the tremendous convergence and divergence of inputs to cortex (part *b* in the figure).

A single cortical neuron receives several thousand synapses, very few of which are from any one presynaptic axon⁵. Similarly, single thalamocortical axons have broad arborizations and contact many hundreds of cortical cells⁶. If the convergence and divergence of intracortical connections is added to the thalamocortical spread, the potential cortical area over which a point of skin can be represented (the anatomical point spread in cortex)



Role of coincident inputs in creating cortical neuron receptive fields. *a*, Normal receptive fields and receptive fields recorded after training. Normally, the receptive field of a neuron in area 3b of primate somatosensory cortex is limited to a single finger (left). Training monkeys in a protocol that caused synchronous stimulation of either the distal or proximal segments of three fingers (middle) caused single neuron receptive fields to span several fingers (right). Not only were skin surfaces that were stimulated together represented together through multiple-finger receptive fields, skin surfaces that were not stimulated in coincident fashion (distal as opposed to proximal segments) were segregated in the cortical representation. *b*, A possible explanation for how coincident inputs create receptive fields. A neuron in cortex (triangle) receives input from a large region of skin; focal regions of skin project anatomically to extensive regions of cortex (middle). The spatiotemporal pattern of skin stimulation determines which connections have greater efficacy (unbroken arrows). Normally, skin regions on a single finger are stimulated together, leading to single-finger receptive fields (left). When monkeys are trained so that several fingers are stimulated together, receptive fields are reorganized orthogonally (right).

to reflect the temporal synchrony of stimulation during training.

Wang *et al.* demonstrate that the changes they describe are due to cortical mechanisms, although others have described changes at subcortical levels of the somatosensory pathway following manipulations of afferent input⁴. Dynamic changes in maps are likely to accompany the learning of spatiotemporal skills by humans as well, but how are such dramatic changes possible in the adult brain?

can be very large indeed. Conversely, the potential area of skin surface that provides input to a cortical cell is extremely large.

A cell's 'receptive field', measured as the region of receptor surface which causes a spike response when stimulated, is a subset of this large skin area. The role of input timing could then be physiologically to increase the effectiveness of a subset of synapses (those that are repetitively coactive), and possibly decrease the effec-

tiveness of others, from the broad set available anatomically. Thus, a cell's receptive field could change, often in dynamic fashion based on the history of stimulation, without changing the anatomical connections that provide input to the cell.

These patterns of anatomical connections to cortical neurons are set up during development and would appear to define the limits of the kind of adult physiological plasticity demonstrated here (though long-term reorganization in the adult cortex following removal or ablation of peripheral input can also involve sprouting of axons into denervated territory⁷). The timing of afferent inputs is crucial for shaping projections not only in the adult cortex, but in the developing cortex as well, as demonstrated by manipulations of visual input during development. Artificially induced strabismus (squint), which alters the temporal synchrony of inputs from the two eyes to visual cortex without altering the overall quantity of input, enhances eye-specific segregation of thalamocortical⁸ and intracortical connections⁹. Routing visual inputs to auditory cortex early in development changes the temporal structure and spatial pattern of input activity to auditory cortex, creating visual receptive fields and maps there¹⁰.

Input correlation-dependent plasticity in cortical responses occurs on a range of timescales¹¹, and some of the mechanisms underlying adaptive changes in the adult brain might share significant features with those in the developing brain. Perhaps the critical difference between the role of input timing in regulating connections in the developing as opposed to the mature cortex is that the physiological changes in synaptic efficacy in developing cortex lead to subsequent anatomical changes, whereas the anatomical substrate remains relatively unaltered in the mature cortex. □

Mriganka Sur is in the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Turning off the water

Russell J. Hemley

THAT the chemistry of materials can be radically different under pressure is one of the essential results of high-pressure research in recent years. Pressure can turn on new chemical reactions, alter kinetics and make possible new materials. Investigations of simple molecular materials have been particularly informative as a result of their simple electronic structure and high compressibility, and the variety of their metastable transformations. And of these systems, those involving water continue to present new surprises. On page 44 of this issue, a study by Loubeyre and LeToullec¹ adds another twist to this theme. They find that the normally explosive reaction between hydrogen and oxygen, surely one of the best studied of chemical reactions, is shut down by pressure at ambient temperature. Moreover, instead of the reaction producing water, they find evidence for a new compound in the system.

The simple combustion of H_2/O_2 mixtures under pressure has been exhaustively studied since the earliest days of chemistry, but details of the reaction kinetics have always been enigmatic. The reaction of interest today is the ideal clean fuel source. Also, the cosmic abundance of these elements is high, rendering the system of great importance to planetary science.

Loubeyre and LeToullec¹ prepared mixtures of these simple gases and loaded them into diamond-anvil cells. During the loading process, the pressure was increased carefully in several steps because at intermediate densities the combustion process occurs readily. (In one inauspicious loading, the pressure was increased too quickly and the mixture ignited.) They used visual observations to determine the phase diagram, together with Raman spectroscopy which provides an ideal means to characterize molecular bonding and can be used to determine the relative concentration of each molecular species within a given phase. Thus, the authors could measure directly the persistence of H_2 and O_2 and the appearance of H_2O , the expected reaction product.

What they found was remarkable: the diatomic molecules persisted to very high pressures and no water formed. The effect of pressure is to slow down the rate of reaction to an astonishing degree. This is unexpected because studies of the gases at lower densities indicate that the reaction rate should increase with pressure. So there must be an unexpected turn-around in the previously documented rate-density relation².

Instead of water, there is evidence for a new, kinetically stable compound in the

system — a phase consisting of intact diatomic molecules with a stoichiometry near $(H_2)_4(O_2)_3$. Such high-pressure 'van der Waals' compounds, or ordered alloys, were first documented in He- N_2 mixtures³ and have been found in a growing number of binary mixtures, but never before in a system as reactive as hydrogen-oxygen. What generally drives the formation of such compounds is packing effects; that is, the molecules fit together more efficiently in the compound than separately as pure phases, so the higher density gives a lower free energy. Such compounds can in principle have rather unusual stoichiometries — for instance, a helium-nitrogen compound where the phase had a formula of $He(N_2)_{11}$. Indeed, surprisingly complex structures have been predicted from simulations of binary mixtures in which the particles are treated as hard spheres⁴. It must be said, however, that identification of the compound in the H_2-O_2 system remains tentative. Spectroscopic properties of the reported compound are very similar to those of the pure phases, and crucial diffraction data are not yet in hand.

The results could have important implications in several areas. First, the new-found pressure control of the $H_2 + O_2$ combustion reaction may be useful for energy technology. Dense, quiescent mixtures of these otherwise reactive molecules might provide a convenient means for fuel storage. Moreover, the shift in the stability of the reported compound to lower densities with decreasing temperature suggests that this (or related) compounds could be exploited as fuel sources.

Beyond this, the results may be relevant to the outer planets, where hydrogen is in abundance. The outer layers of the large gaseous planets are believed to contain clouds of solid condensates at different depths. Could metastable H_2-O_2 compounds form within these layers? It is likely that the effects of increasing temperature with depth and heterogeneous reaction kinetics would drive combustion to completion, but this remains to be determined by quantitative measurement of the reaction kinetics. Loubeyre and LeToullec did unexpectedly obtain some information on the kinetics during an attempt to determine the structure by X-ray diffraction with an intense synchrotron beam. This resulted in decomposition and the formation of H_2O , confirming that water is still the stable phase and, furthermore, that the reaction is not completely shut down by pressure: heating, as yet unquantified, can drive the reaction towards the stable product. ▶

1. Wang, X., Merzenich, M. M., Sameshima, K. & Jenkins, W. M. *Nature* **378**, 71–75 (1995).
2. Bonhoeffer, T. & Grinvald, A. *J. Neurosci.* **13**, 4157–4180 (1993).
3. Merzenich, M. M. et al. *Neuroscience* **8**, 33–55 (1983).
4. Nicolais, M. A. L., Lin, C. S., Woodward, D. J. & Chapin, J. S. *Nature* **361**, 533–536 (1993).
5. Peters, A., Payne, B. R. & Budd, J. *Cerebral Cortex* **4**, 215–229 (1994).
6. Rausell, E. & Jones, E. G. *J. Neurosci.* **15**, 4270–4288 (1995).
7. Darian-Smith, C. & Gilbert, C. D. *Nature* **368**, 737–740 (1994).
8. Shatz, C. J., Lindstrom, S. & Wiesel, T. N. *Brain Res.* **131**, 103–116 (1977).
9. Lowel, S. & Singer, W. *Science* **255**, 209–212 (1992).
10. Sur, M., Pallas, S. S. & Roe, A. W. *Trends Neurosci.* **13**, 227–233 (1990).
11. Armstrong-James, M., Diamond, M. E. & Ebner, F. F. *J. Neurosci.* **14**, 6978–6991 (1994).

Perhaps the greatest implications are for fundamental chemistry. The study of H₂O under pressure continues to reveal new phenomena. The discovery over a decade ago⁵ of pressure-induced amorphization of ice Ih and different amorphs of H₂O came as a considerable surprise and is the subject of continuing investigations. Experiments on mixtures of H₂ and H₂O reveal new classes of dense clathrates⁶, including a high-density H₂-H₂O clathrate in which diatomic hydrogen is preserved to pressures in excess of 50 GPa. The O-H bonds are not easily severed by pressure; indeed, H₂O ice is stable at over one megabar (at least 128 GPa)⁷. But even below this pressure, spectroscopic data indicate that the solid is no longer truly molecular — the hydrogen bonds are approaching a symmetrized state⁸ and the material should be more properly considered an oxide. The comparison is instructive: the work done in compression (pressure-volume work) is comparable to the chemical bond-

strengths and thus is sufficient for rearranging chemical bonds. So the creation of new solids — both stable and metastable — in the H₂-O₂ binary system at still higher pressure remains an exciting challenge for high-pressure chemistry. Meanwhile, the observations of Loubeyre and LeToullec have added another intriguing chapter to the metastable behaviour of this system. □

Russell J. Hemley is at the Geophysical Laboratory and Center for High-Pressure Research, Carnegie Institution of Washington, Washington DC 20015, USA.

1. Loubeyre, P. & LeToullec, R. *Nature* **378**, 44–46 (1995).
2. Astapov, N. S. *et al. Fiz. Goreniya Vzryva* **20**, 98–105 (1984).
3. Vos, W. L. *et al. Nature* **358**, 46–48 (1992).
4. Elkridge, M. D., Madden, P. A. & Frenkel, D. *Nature* **365**, 35–37 (1993).
5. Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **310**, 393–394 (1984).
6. Vos, W. L., Finger, L. W., Hemley, R. J. & Mao, H. K. *Phys. Rev. Lett.* **71**, 3150–3153 (1993).
7. Hemley, R. J. *et al. Nature* **330**, 737–740 (1987).
8. Pruzzan, P. *J. molec. Struct.* **322**, 279–286 (1994).

SEXUALITY

Another important organ

S. Marc Breedlove

THE view of the brain as a sexual organ gains further credibility from a report on page 68 of this issue by Zhou *et al.*¹, who show that there is a structural difference in the brains of 'ordinary' men and those of a group of men who wanted to become women (male-to-female transsexuals). The authors emphasize that fetal hormone levels might account for these brain differences, so this paper is sure to fuel debate about whether human sexual behaviour is shaped primarily by society or by biological variables. But the difficulties inherent in studying the diverse sexual behaviour of humans ensure that this will be far from the final word on the subject.

The tiny region of the brain that is under scrutiny is the central subdivision of the bed nucleus of the stria terminalis (BSTc), and is part of the hypothalamus, which helps to keep the different systems in the body working in harmony. The BSTc is defined on the basis of innervation by the neurotransmitter VIP (vasoactive intestinal polypeptide), and has been implicated in sexual behaviour in non-human animals. Post-mortem analysis revealed that the volume of this brain region is smaller in women than in men and that the six transsexual men also had a small BSTc. Thus men who wished to become women had a smaller, more 'feminine' BSTc than other men.

These observations are reminiscent of an earlier reported difference in the brains of heterosexual and homosexual men, in a nucleus known as INAH-3 in

the preoptic area². Both relate human brain structure to psychological sexual differentiation. But the difference discovered in the BSTc size is clearly not related to sexual orientation. First, male-to-female transsexuals vary considerably — some are attracted to men, some to women, some are bisexual and some have no interest in sexual relations at all. Transsexuality does not seem to be so much a phenomenon of sexual orientation as of body image. Second, a cohort of brains from homosexual men revealed BSTc volumes as large as those from the brains of presumed heterosexual men.

There are two reservations that might undermine the correlation described by Zhou *et al.*¹. The first is that some other feature, more mundane than the fascinating and psychologically complex phenomenon of transsexuality, is in fact responsible for the correlation between BSTc size and the transsexuality of the subjects. Within the limitations of working with post-mortem measurements from human subjects, Zhou *et al.* tried to address this possibility. For example, all of the transsexuals had received oestrogen treatments, all had taken the antiandrogenic drug cyproterone acetate, and all but one had been castrated. So the transsexuals, like most women, had low androgen and high oestrogen concentrations in their blood. Perhaps these hormone conditions altered the expression of VIP, which in turn affected BSTc sizes.

Steroid manipulation in adults certainly

alters neuropeptide expression in non-human animals³. Indeed, the two transsexuals with the largest BSTc (but still within the women's size range) were those who had stopped taking oestrogen before they died. There were several other non-transsexual individuals in the study who, because of various medical conditions, had low androgen or high oestrogen concentrations before death, but the BSTc remained small in the women and large in the men.

Of course such anecdotal comparisons do not rule out alternative explanations. However, humans differ from one another in so many respects that it is not practical to assemble large enough groups to control for every variable, particularly when dealing with so rare a condition as transsexuality. So it is reassuring that the examples described contradict any simple relation between confounding variables and BSTc volume, suggesting that it is transsexuality *per se* that co-varies with BSTc size.

The second reservation concerns the ontogeny of the BSTc and represents that chicken-or-the-egg dilemma endemic to brain correlates of behaviour. Did the development of a small BSTc (under the influence of fetal hormone concentrations or gene action alone) cause these males to feel that they were 'women trapped in men's bodies' and seek sex reassignment, or did the development of this body image discrepancy (under the influence of social or other experiences) promote underdevelopment of the BSTc? For that matter, did the transsexuals begin puberty with a large, masculine BSTc, which then shrank in response to the myriad experiences associated with transsexuality?

Laymen may assume that a structural difference in the brain is the immutable signature of purely biological forces, but three decades of neuroscience research have made it clear that experience can dramatically alter the structure and function of the brain^{4–6}. The same laboratory now reporting the differences in the brains of transsexuals was also the first to report a sex difference in the human preoptic area⁷, a region they named the sexually dimorphic nucleus of the preoptic area (SDN-POA), after a similar nucleus found in the rat brain⁸. The SDN-POA was larger in adult men than in adult

1. Zhou, J.-N., Hofman, M. A., Gooren, L. J. G. & Swaab, D. F. *Nature* **378**, 68–70 (1995).
2. LeVay, S. *Science* **253**, 1034–1037 (1991).
3. De Vries, G. J. *J. Neuroendocrinol.* **20**, 1–13 (1990).
4. Bennet, E. L., Diamond, M. L., Krech, D. & Rosenzweig, M. R. *Science* **146**, 610–619 (1964).
5. Hubel, D. H. & Wiesel, T. N. *J. Neurophysiol.* **28**, 1041–1059 (1965).
6. Yang, T. T. *et al. Nature* **368**, 592–593 (1994).
7. Swaab, D. F. & Fliers, E. A. *Science* **228**, 1112–1115 (1985).
8. Gorski, R. A., Harlan, R. E., Jacobson, C. D., Shyrne, J. E. & Southam, A. M. *J. comp. Neurol.* **193**, 529–539 (1980).
9. Swaab, D. F., Gooren, L. J. G. & Hofman, M. A. *Progr. Brain Res.* **93**, 205–219 (1992).

women. But subsequently this group reported that there was no sex difference in the numbers of neurons in the human SDN-POA until ten years of age⁹.

Thus there is ample room for early experiences, including the very different social stimuli presented to boys and girls, to cause the sex difference found in the adult SDN-POA. Similarly, we do not know how INAH-3 develops in humans, so it is not clear whether this sexual dimorphism is a cause or an effect of human sexual differentiation. Neither do we know whether some men having a small INAH-3 at birth are therefore born gay or, more correctly, born to become gay as adults. Only a sexual dimorphism found in newborn or fetal humans could safely be described as unrelated to social stimulation.

At present, these brain regions can only be measured post mortem. Until technology enables us to measure them

repeatedly in the same person at different ages (before and after puberty, for example), there can be no definitive answer as to whether these regions direct psychological sexual differentiation or are themselves directed by that process.

Of course, these alternatives are not mutually exclusive. It may be that some are born with a predisposition to transsexuality and that social influences either potentiate or attenuate this predisposition. Both predisposition and social influence would have to affect the brain physically and the work of Zhou and colleagues suggests that the BSTc may be the locus of such effects. □

S. Marc Breedlove is in the Psychology Department and Graduate Group in Neurobiology, 3210 Tolman Hall, University of California, Berkeley, California 94720-1650, USA.

reconstruct the missing body parts in an orderly way. Likewise, attempts to dissect seedling root development failed genetically but resulted in the isolation of mutants that altered the radial pattern of the body axis both in the embryo and post-embryonically^{5,6}. Interestingly, these mutants displayed the same pattern defects in lateral roots as well as in 'adventitious' roots regenerated from lumps of undifferentiated cells called callus, although both kinds of root have a non-embryonic origin.

The formation of lateral roots might shed some light on the role of the meristem in development. Initially, a small group of pericycle cells starts dividing, producing a lateral root primordium which subsequently forms at its tip a new meristem that is essentially organized like the primary root meristem of the seedling². If radial patterning occurs in the young lateral root primordium, the tissue pattern thus established might serve as a reference for the subsequent extension of the cell files by the root meristem.

Has the paradigm been turned around, making the former master of development its slave? The answer is, not really. The root meristem is not just a cluster of proliferating cells, supplying the plant body with raw material for continued growth. The root meristem has a specific organization, with its stem cells being arranged around a quiescent centre. This organization is restored following ablation of the quiescent centre³, or excision of the quiescent centre and the root cap⁷, suggesting that this central group of cells plays an as-yet undefined role in the functional organization of the root meristem⁸. One possible role of the quiescent centre could be to keep the physically adjacent stem cells dividing.

Unexpected results often raise more questions than can be answered. One of the most challenging tasks will be to determine the nature and origin of the root patterning signal(s). In view of a recent report on the cell wall and the plant extracellular matrix being involved in cell-fate determination in the brown alga *Fucus*⁹, we should be prepared to face more surprises in plant development. □

Gerd Jürgens is in the Department of Developmental Genetics, University of Tübingen, Spemannstrasse 37-39, D72076 Tübingen, Germany.

DEVELOPMENTAL BIOLOGY

Rooting the meristem

Gerd Jürgens

PLANT developmental biology was focused for a long time on groups of stem cells, the meristems of the shoot and root, that give rise to adult structures during post-embryonic development. A large body of experimental evidence led to the traditional view that meristems are autonomous pattern-generating machines¹, producing the right kinds of organs and cell types at the right place. It seems ironical that the meristem with the most stereotyped arrangement of stem cells, the root meristem of the *Arabidopsis* seedling², has now failed the critical test. As reported by Scheres and colleagues on page 62 of this issue³, laser ablation of specific cells changes the fate of adjacent cells, indicating that the stem cells of the root meristem lack intrinsic pattern-generating information.

The root meristem is organized around a core of mitotically inactive cells termed the quiescent centre, which splits the stem cells into an upper and a lower tier of initial cells. While the lower tier adds new layers to the root cap, thereby shielding the root tip against mechanical destruction, the upper tier gives off daughter cells that extend the tissue-specific cell columns known as files which run the length of the growing root². In cross-sections of the root, the tissues are arranged in a radial pattern like the rings of a tree: the vascular tissue in the centre is surrounded by concentric layers of pericycle, endodermis, cortex and epidermis cells. Not only is the radial sequence of tissue layers fixed, but the number of cell files within each tissue is also invariant, or nearly so. For example, there are

always eight files each of cortex and endodermis cells derived from eight common stem cells which continually produce daughter cells that divide periclinally (radially) to give an outer cortex and an inner endodermis cell.

In their elegant series of laser-ablation experiments, van den Berg *et al.*³ destroyed either stem cells or their daughter cells in order to determine the origin of cortex/endodermis cell fate. If an individual stem cell is ablated, the underlying pericycle stem cell divides abnormally and its outer daughter cell effectively takes over the function of a cortex/endodermis stem cell. If a single daughter cell is destroyed, the cortex/endodermis stem cell continues to function normally. But if three adjacent daughter cells are ablated, the central one of the three stem cells now produces daughter cells that fail to divide properly to give cortex and endodermis cells. If instead a radial set of three daughter cells, one each from epidermis, cortex/endodermis and pericycle, is eliminated, the cortex/endodermis stem cell continues to function normally. The authors conclude that the patterning information does not reside in the stem cells of the meristem themselves, but seems rather to be imparted by mature cells of the root.

In retrospect, the present results extend earlier observations made using different approaches. For example, basal amputation of older carrot somatic embryos resulted in proximo-distal regeneration of root and root meristem from the cut edge of the stump⁴, indicating that the remaining mature tissue has the information to

1. Steeves, T. A. & Sussex, I. M. *Patterns in Plant Development* (Cambridge University Press, 1989).
2. Dolan, L. *et al. Development* **119**, 71-84 (1993).
3. van den Berg, C., Willemsen, V., Hage, W., Weisbeek, P. & Scheres, B. *Nature* **378**, 62-65 (1995).
4. Schiavone, F. M. & Racusen, R. H. *Development* **113**, 1305-1313 (1991).
5. Benfey, P. N. *et al. Development* **119**, 57-70 (1993).
6. Scheres, B. *et al. Development* **121**, 53-62 (1995).
7. Feldman, L. J. *Planta* **128**, 207-212 (1976).
8. Kerk, N. M. & Feldman, L. J. *Development* **121**, 2825-2833 (1995).
9. Berger, F., Taylor, A. R. & Brownlee, C. *Science* **263**, 1421-1423 (1994).

Diamonds everywhere

Christian Koeberl

DIAMONDS have long caused excitement not only among the general public, which appreciates them mainly for their beauty and value, but also among geologists. The latter use diamonds in a not very glamorous way as probes of the properties of the Earth's mantle, which is otherwise rather inaccessible. On Earth, diamonds usually occur in rocks derived from the Earth's mantle — igneous rocks such as kimberlites, or metamorphic rocks of mantle origin — and are thought to have formed from fluids or melts in the upper mantle at immense pressures and temperatures.

Considering the copious information available on the 'common' mantle-derived diamonds, many may be surprised to find out that diamonds have also formed in nature directly on the Earth's surface. On page 41 of this issue, Hough *et al.*¹ add an important discovery to the small body of research on diamonds produced during meteorite impacts on Earth. In their studies of material from the 23-km-diameter Ries crater in Germany, they found minute quantities not only of cubic and hexagonal diamond (the latter is known as lonsdaleite), but also cubic diamond intergrown with silicon carbide. Never before

have such composites of diamond and SiC been found in nature. What is more, the diamonds seem to have formed not through shock but from the vapour phase.

Although mantle-derived diamonds probably originated during several diamond-forming events early in the Earth's history, two other natural processes are capable of producing diamonds at any time. Meteoritists have long been aware of the occurrence of diamonds in iron meteorites^{2,3} and ureilites⁴, but initially assumed that the diamonds originated at high gravitational pressures within meteorite parent bodies, as for kimberlitic diamonds on Earth. Later work⁵ indicated that these meteoritic diamonds have formed not in an environment of steady high pressure, but by shock from graphite or amorphous carbon in a mere instant. However, with evidence that the shock event during which these diamonds formed was related to collision and break-up of the meteorite parent body, and not to the impact of the meteorite on the Earth, attention was diverted away from impact craters.

New attention was drawn to diamonds in meteorites when an elusive anomalous noble-gas bearing component in chondrit-

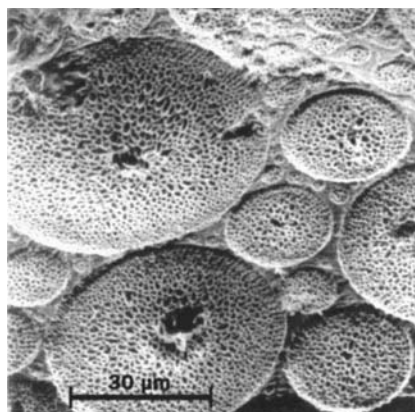
ic meteorites was identified as consisting of nanometre-sized diamonds⁶. Once meteoritists started looking, they found such small diamond grains in surprisingly large quantities (up to several hundred parts per million) in a variety of meteorite types. What is more, the rare-gas composition of these minute diamond grains gave them away as being of interstellar (presolar) origin, having probably formed by vapour condensation in stellar atmospheres. The discovery of diamonds in clays marking the boundary between the Cretaceous and Tertiary periods (K/T boundary)⁷, which contains fairly unambiguous evidence of a large-scale impact, provided the next surprise. The nanometre-sized diamond grains were first interpreted as being remnants of the original meteorite⁷, but carbon and nitrogen isotopic data pointed to an origin during the impact or from within the fireball⁸.

Speculation was rife about the origin of these clearly impact-associated diamonds. Here, earlier work by Soviet scientists (largely unknown outside the Soviet Union, as some aspects of this research were considered to be a state secret) gained importance. Polycrystalline diamonds of up to about 1 cm in size were discovered in impactites at a few Russian and Ukrainian impact structures⁹. These diamonds, which usually contain appreciable amounts of lonsdaleite, have crustal chemical and isotopic signatures¹⁰ and pre-

The natural way to leave an impression

"Every now and then", said D'Arcy Thompson in 1917, "we come to deep-seated signs of symmetry which seem to lie beyond the reach of the ordinary physical forces". This statement came at the conclusion of an exposition on radiolarian skeletons, whose astonishingly symmetrical silica frameworks embellished several pages of Thompson's classic *On Growth and Form*. But Thompson's aim was to demonstrate that "it by no means follows that the forces in question are not essentially physical forces". Geoffrey Ozin and colleagues now seem to have shown the truth of that supposition experimentally on page 47 of this issue, where they report that intricately patterned inorganic materials like the one pictured here can be prepared by what one could justifiably call 'bucket chemistry'.

The formula sounds alarmingly simple: throw together phosphoric acid and pseudo-boehmite (a hydrated aluminium oxide) in tetraethylene glycol in the presence of an alkylamine; heat, dry and recrystallize. The millimetre-sized aluminophosphate spheres that are produced have surface patterns that resemble nothing so much as the silica



skeletons of radiolaria, patterned with arrays of disks, pores and bowls of micrometre dimensions.

Where do these patterns come from? Ozin *et al.* suppose that they are imprinted by vesicles formed by self-assembly of the alkylamine amphiphiles, whose close-packing on the surface of the growing aluminophosphate spheres leaves regularly spaced imprints. The mechanism remains speculative, but it is nothing more than the natural extension of templated patterning of inorganic materials on the molecular and nano-

metre scales. At the former scale there is the synthesis of microporous molecular sieves (synthetic zeolites) from the polymerization of silicate ions in the presence of quaternary ammonium ions, which act as templates for the material's ångström-scale cavities. The latter size regime pertains to the synthesis of ordered mesoporous solids (see *Nature* 359, 710; 1992), where the templates seem to be micelles.

Templating by amphiphilic vesicles would therefore extend this principle to the next rung of the hierarchical ladder, carrying the patterning to the micrometre scale. It might also provide a model system for investigating how radiolaria themselves perform their delicate handiwork. Pleasingly, this would not only broaden the emerging picture of pattern formation in materials via organic-inorganic interactions but could establish that D'Arcy Thompson, by proposing that "the peculiar form and character of these little skeletons are due to the moulding of [inorganic] material upon an underlying vesicular structure", was demonstrating another example of the prescience with which his book is filled.

P. B.

serve the macroscopic crystal habit of the graphite crystals from which they formed by shock metamorphism. The characteristics of the impact-produced diamonds, which are likely to have formed by martensitic transformation of primary graphite, are quite different from those of microdiamonds occasionally found in minerals of high-grade metamorphic rocks¹¹.

The nanodiamonds at the Ries crater are unlikely to have formed by the same mechanism as the larger polycrystalline impact diamonds, which also contain evidence of high-pressure gas inclusions¹⁰. The morphology of the nanodiamond aggregates and the SiC intergrowths¹ seems to be in better agreement with an origin by condensation (chemical vapour deposition, well known to materials scientists as CVD) in the hot vapour cloud of the impact. This is somewhat similar to the origin of the nanodiamonds found within stony meteorites, only the latter have condensed in stellar atmospheres. The Ries studies also support evidence that the K/T boundary nanodiamonds formed during the impact process, most probably also in a CVD-like process, and are not remnants of the meteorite itself. Further evidence may result from a search for SiC in K/T boundary sediments.

There is, however, another rare and unusual form of diamond on the Earth. Polycrystalline black diamonds (known as 'carbonado' from the Portuguese for 'carbonated') occur within sediments in Bahia (Brazil), and Ubangi (Central African Republic). Carbonados are extremely hard polycrystalline aggregates with a porous matrix of randomly oriented crystallites (0.5–20 µm in size), which often enclose octahedral and cubic diamonds and other noncarbonaceous inclusions. No carbonado diamond has ever been found *in situ* in a rock.

The carbonados are another can of worms, as no one knows for sure how they formed. In fact, precious few detailed studies are available, and those that exist have led to such diverse suggestions for their origin as carbon subduction in the mantle, shock metamorphism during impact, and irradiation of organic matter¹². In view of the crustal signatures seen in carbonados, a mantle origin is unlikely,

but the jury is still out on a possible impact connection. Many open questions remain. For example, if carbonados are the result of Precambrian impacts, what was the source of the carbon? Did they form by condensation or by shock metamorphism? Is there any independent evidence for such a large impact event? Detailed isotopic and structural work, as in the case of the Ries diamonds, may help to settle their origin.

One conclusion becomes increasingly

obvious. There are more forms of diamonds, and ways of making them, between heaven and Earth than were dreamt of a few years ago. Large-scale impact processes have been good for surprises before. The new diamond studies indicate that yet another revelation may be around the corner. □

Christian Koeberl is in the Institute of Geochemistry, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria.

METEORITICS

Parental paradoxes

Monica M. Grady

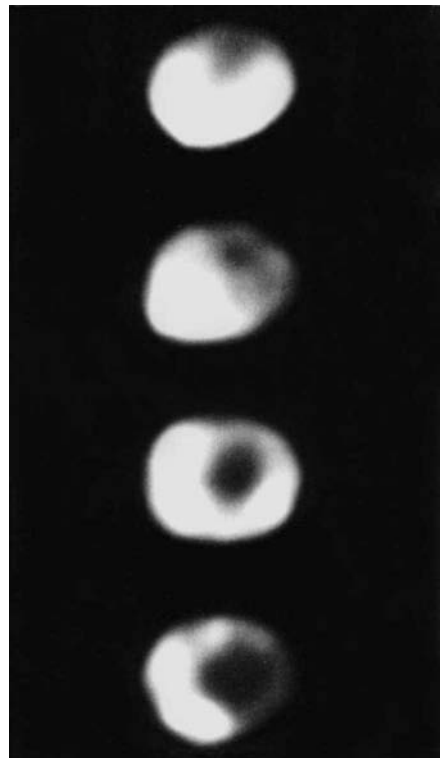
ALMOST all meteorites come from the asteroid belt, a zone of objects in near-circular orbits between Mars and Jupiter. This belt is not simply a homogeneous assembly of material left over from the formation of the Solar System, but contains a diversity of material distributed across a wide range of orbits. Just as meteorites can be classified into groups on the basis of mineralogy and chemistry, so too can asteroids be grouped on the basis of their spectral reflectance patterns. But how valid is it to assign specific groups of meteorites to specific classes of asteroid parents on the basis of spectroscopic data alone? A special session at a recent conference* generated a typically lively debate between astronomers (who observe asteroids) and meteoriticists (who analyse meteorites) over the interpretation of spectral data and subsequent assignation of meteorite group to asteroid parent. The debate looks set to continue until an asteroidal fragment captured in space is returned to Earth.

There are two separate puzzles that have yet to be solved concerning the origins of the two largest groups of stony meteorites. For one group, the ordinary chondrites, there are many meteorites but no associated parent asteroid type. In contrast, the parent asteroid that produced the eucrite meteorites should be completely fragmented, but is apparently still present as the asteroid 4 Vesta.

The ordinary chondrites—thought to be primitive meteorites little changed since the birth of the Solar System—are, as their name implies, the most abundant type of meteorite observed to fall on Earth. By inference, then, it might be assumed that ordinary chondrite parents would dominate within the asteroid belt. This is far from the case. Although one very small asteroid (3628 Boznemcova) has been identified as having a reflectance spectrum similar to that of ordinary chon-

drites (R. Binzel, Massachusetts Inst. Technology), this single body is insufficient to account for these abundant meteorites.

Recent results from observations made by the spacecraft Galileo of the S-type asteroids Gaspra and Ida, and Ida's small moon Dactyl, might help explain the dichotomy between observed and recovered species. Cratering, it seems, has altered the surfaces of Gaspra and Ida, by production of a deep soil layer, or regolith, in such a way that 'space weathering' has modified the spectral reflectance patterns (C. R. Chapman, Univ. Arizona). The tiny moon Dactyl, which probably has the least overburden of regolith, and thus has suffered the least from space weathering, has a reflectance



Asteroid 4 Vesta, pictured by the Hubble Space Telescope at intervals throughout its 5.34-hour rotation.

B. Zellner (Georgia Southern Univ./NASA)

1. Hough, R. M. et al. *Nature* **378**, 41–44 (1995).
2. Weinschenk, E. *Ann. naturhistor. Hofmuseums, Wien*, **4**, 93–101 (1889).
3. Foote, A. E. *Am. J. Sci.* **42**, 413–417 (1891).
4. Urey, H. C., Mele, A. & Mayeda, T. *Geochim. cosmochim. Acta* **13**, 1–4 (1957).
5. Lipschutz, M. E. & Anders, E. *Science* **134**, 2095–2099 (1961).
6. Lewis, R. S., Ming, T., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
7. Carlisle, D. B. & Braman, D. R. *Nature* **352**, 708–709 (1991).
8. Gilmour, I. et al. *Science* **258**, 1624–1626 (1992).
9. Masaitis, V. *Meteoritics* **27**, 21–27 (1992).
10. Koeberl, C. et al. *Lunar planet. Sci.* **XXVII**, 777–778 (1995).
11. Sobolev, N. V. & Shatsky, V. S. *Nature* **343**, 742–745 (1990).
12. Ozima, M., Zashu, S., Tomoura, K. & Matsuhisa, Y. *Nature* **351**, 472–474 (1991).

*58th Meeting of the Meteoritical Society, Washington DC, USA, 11–15 September 1995.

spectrum closest to that of the ordinary chondrites. So, by implication, the S-type asteroids (or a subset thereof) might be the space-weathered parents of ordinary chondrites. An extension of this observation is the paradox that although meteorites (including ordinary chondrites) are derived from main-belt asteroids, these asteroids are so cratered and weathered that they may bear little resemblance to the original planetesimals that aggregated at the time of Solar System formation (Chapman). Even this solution is not clear-cut: the effects of space weathering are not fully understood, and it was also argued that shock-blackened C-type asteroids might be the parents of ordinary chondrites (D. T. Britt, Univ. Arizona).

In contrast to the ordinary chondrites, the eucrite assemblage of meteorites has long been believed to originate from the large, bright asteroid 4 Vesta. Eucrites are differentiated meteorites, basaltic entities formed by melting on a parent body, and their reflectance spectra are remarkably similar to that of Vesta. Notwithstanding the observational data, there are still problems associated with this asteroid-meteorite pairing. J. T. Wasson (Univ. California, Los Angeles) reminded us that the eucrites, which are crustal material, have the same oxygen isotopic composition, and thus are closely associated with two other classes of meteorite that must derive from the core and core-mantle boundary of a parent body (the IIIAB iron meteorites and the pallasites, respectively). On these grounds, the parent asteroid must have been destroyed, and cannot be Vesta.

Others were convinced that irrefutable links had been forged between Vesta and the eucrites (Binzel). The first link in the chain was the close match in reflectance spectra between the asteroid and the meteorites, a strong link identified many years ago. The second link, escape from Vesta and subsequent delivery to the Earth, has only recently been fashioned. Asteroids leave the asteroid belt through certain, well defined 'escape hatches', governed by Jupiter's presence. One of the drawbacks to Vesta's identification as the eucrite parent was its distance from such a position. This dynamic delivery difficulty has been overcome by the observation of a swarm of asteroids with spectroscopic signatures close to that of Vesta. The attendants occupy the region of space between Vesta and the 3:1 jovian resonance 'escape hatch', thus providing both a route by which Vesta-like material could leave the asteroid belt and make its way to Earth, and the second link in the chain. The final connection between Vesta and the eucrites is the successful delivery of material to Earth, as recognized by the presence of the eucrites themselves — small pieces of an asteroid observed to fall and preserved as meteorites (see W. B.

McKinnon, *Nature News and Views* **263**, 211–212; 1993). The division between astronomers and meteoriticists in this case becomes more explicable if the assemblage of eucrites, pallasites and IIIAB iron meteorites is not from a single body, but from several bodies sharing a similar reservoir of material (R. N. Clayton, Univ. Chicago). The attendant swarm of small asteroids might provide the parents of pallasites and IIIAB irons. But if this were the case, a significant fraction of potential material is still missing — we have core, core-mantle boundary and crust, but no mantle material from the parents represented in our meteorite collections. The missing mantle is a puzzle still requiring solution.

It was apparent that, although recent images from Galileo and the Hubble Space Telescope have provided much valuable new information about asteroids, astronomers and meteoriticists are still left facing paradoxes when trying to assign meteorites to parents. Plausible as the

explanations of space weathering, shock blackening and asteroid swarms were, opinion at the meeting still remained divided over parental assignments. Outstanding questions concerning the absence of crust and mantle material from the 70 or so iron meteorite parents, missing mantle from the eucrite parent and the identification of only small asteroids as possible ordinary chondrite parents were also raised and discussed. Much hope is being pinned on the forthcoming NASA asteroid rendezvous mission (NEAR, launch date February 1996), to yield further spectroscopic data which might help link the 'ground truth' of meteorites with their enigmatic asteroidal parents. Meanwhile, astronomers and meteoriticists will continue to debate, if not until the cows come home, at least until the data are returned. □

Monica M. Grady is at the Natural History Museum, Cromwell Road, London SW7 5BD, UK.

GEOPHYSICS

A wayward plume?

Norman H. Sleep

PLATE tectonics accounts neatly for the commonest types of volcanism on the Earth — the mid-ocean ridges that produce new ocean crust, and the chains of volcanoes that lie behind deep-sea trenches where one plate is subducted beneath another. But what of the more localized volcanism that either amplifies the output of a mid-ocean ridge (as in Iceland) or occurs away from the plate boundaries (as in Hawaii)? Both types of 'hotspot' volcanism are generally thought to have the same underlying cause: a hot plume of material upwelling from the mantle. Such features begin as a buoyant spherical head fed by a tubular conduit (the tail) formed by low-viscosity hot mantle. The arrival of a plume head at the base of the lithosphere results in uplift, eruption of flood basalts and sometimes continental break-up.

Despite the appeal of the theory, confirmation has proved difficult, and direct detection of either plume heads or tail conduits is of great interest to the earth sciences. Two years ago, Nataf and VanDecar reported a possible detection of a plume upwelling under the ocean floor¹. Now, on page 25 of this issue, VanDecar *et al.* report that they may have found the fossil traces of a plume beneath a continent².

VanDecar and co-workers have carried out a seismic tomography study which finds a region of low seismic velocity in the upper mantle beneath the Brazilian shield, at a depth of 200 to 500 km. Low seismic velocity generally implies high temperatures in the rock, and they

associate this region with the plume head that produced the Paraná basalts in the region around 135 million years ago. If so, it must be an ancient relic indeed. As the authors note, a plume is not expected there from conventional theory, for, in the intervening millennia, the surface plate has migrated thousands of kilometres west of the deep mantle source of the Paraná plume, which is alive and well beneath the Tristan hotspot. They conclude that the 250-km-wide low-velocity region was heated by the passage of the plume head through upper mantle that has since remained fixed with respect to the overlying South American plate.

For this interpretation to be correct, the hot region must still be about 200 K hotter than normal mantle, even though it is no longer being supplied by hot material through the tail conduit. As VanDecar *et al.* note, an original 200-km-wide region with a temperature contrast of 400 K could have evolved to the current region over the past 135 million years by thermal diffusion.

Flow within the mantle is a more serious enemy of local hot regions. First, a hot region is expected to ascend buoyantly towards the base of the lithosphere. The ascent rate (30 kilometres per million years) estimated from Stokes law by taking into account the radius of the buoyant region, its density contrast, the acceleration of gravity, and the expected viscosity of the surrounding mantle, is an order of magnitude too rapid for the region to persist in the upper mantle for 135 million

years. Thus the anomalous region needs to have either an unusually low density contrast with its surrounding for its seismic velocity anomaly or be within upper mantle material that is an order of magnitude more viscous than typical estimates of upper mantle viscosity.

Second, the westward motion of the South American plate relative to the deep mantle would shear the anomaly. In this case, about 3,000 km of motion between the plate and the lower mantle has occurred over a 600 km depth range. If this shear was evenly distributed with depth, an original 250-km-diameter sphere would have been smeared out over 1,250 km. For the sphere to survive, shear would have to be accommodated by an underlying low-viscosity layer.

Given these restrictions, could the low-velocity anomaly instead be associated with a more recently formed hot region? Unfortunately such regions are not expected from conventional theory. Another possibility is a normal-temperature region surrounded by cool downwellings associated with convection at the base of the lithosphere — as tomography detects only lateral viscosity contrasts, this would not be distinguished from a hot upwelling. But cool downwellings are expected to have small temperature contrasts (enough to change viscosity by a factor of e)^{3,4}.

A new starting plume would also need to have a small head — if there were a large head, we would expect to see volcanism, uplift and a much broader seismic anomaly today. But it is difficult to create a small starting plume head because head radius is proportional to the one-fifth power of volume flux of the plume^{5,6}. In addition, starting plume heads spend most of their life in the deep high-viscosity parts of the mantle and thus should be infrequent at shallower depths. For example, if viscosity increases exponentially with depth, the derivation in ref. 5 is easily modified to show that 76 per cent of the ascent time is spent in the basal layer where viscosity is within a factor of e of its maximum value. The upward decrease of viscosity in the Earth may generate small plume heads but they are expected to reach the surface soon after the primary one⁶. Therefore a convective explanation for the anomaly would have to appeal to poorly understood processes, perhaps involving the phase change at 670 km depth or the tail of a weak plume that is not associated with a surface hotspot.

Could the anomaly be an artefact of the tomographic analysis? VanDecar *et al.*² have been careful in their analysis. They have correlated wave forms to pick arrival times to a precision of 0.03 s. Their successful model reduces the variance of residuals to this level. This avoids the 1-s data precision, systematic errors and slight though statistically significant reductions in variance of arrival times compiled in

bulletins. They have also done a resolution analysis to show that their data would detect an anomaly of this size if it were present. At a minimum, their velocity anomaly is a viable explanation of their observations.

Undersampling of the Earth by seismic waves is an unavoidable problem in the study. Seismic waves sample a region of about a wavelength, here 8 km, around their ray paths and the distribution of sources and receivers is such that many 8-km cubes are missed altogether by ray paths. Thus it is necessary to have a much coarser model resolution, about 100 km. This is obtained by a smoothness constraint in the least-squares inversion. The alternative of using 100-km waves which do not undersample is not feasible because the arrival times cannot be accurately picked.

The smoothness constraint (or alternatively a constraint on the root-mean-square value of model perturbations) is necessary to obtain a stable inversion, but the cost is far greater than just defocusing the resolved anomaly. The amplitude of the anomaly times volume may be increased (or decreased), real strong anomalies may be suppressed and strong artefact anomalies may be created in poorly sampled regions far from the real anomalies.

In the case of downgoing slabs, strong velocity anomalies are known to exist. The common practice of using spatially uniform smoothing (or perturbation) constraints is tantamount to assuming that nothing is known about plate tectonics; it suppresses the real slab anomaly and causes artefacts elsewhere. However, no upper-mantle anomaly is expected beneath the Brazilian shield so a simple smoothing constraint is all that can be applied. A scheme for producing the observed anomaly as an artefact of strong anomalies elsewhere is certainly not evident.

In conclusion, the results of VanDecar *et al.* are puzzling but cannot be easily dismissed. Either a hot region is associated with the observed anomaly, or compositional variations or anisotropy produce velocity anomalies not associated with density. In the former case, conventional plume theory needs to be modified. In the latter, a caveat is posed for studies (such as ref. 1) that attempt to detect plume tails beneath hotspots. □

Norman H. Sleep is in the Department of Geophysics, Stanford University, Stanford, California 94305, USA.

1. Nataf, H.-C. & VanDecar, J. C. *Nature* **364**, 115–119 (1993).
2. VanDecar, J. C., James, D. E. & Assumpção, M. *Nature* **378**, 25–31 (1995).
3. Davaille, A. & Jaupart, C. *J. Fluid Mech.* **253**, 141–166 (1993).
4. Davaille, A. & Jaupart, C. *Geophys. Res. Lett.* **320**, 1827–1830 (1993).
5. Olson, P. in *Magma Transport and Storage* (ed. Ryan, M. P.) 33–51 (Wiley, New York, 1990).
6. Bercowski, D. & Mahoney, J. *Science* **266**, 1367–1369 (1994).

Spectral shifts

COLOUR matching is a tricky visual skill, as anyone knows who has had to paint in a chip or a scratch on a car body with a matching shade. Daedalus is now inventing a single universal paint to match all possible cars.

He points out that a colour photograph is slowly and selectively decolorized by daylight. The blue and ultraviolet rays bleach out all the colours that absorb them. The result is a monochrome image in the one colour that doesn't absorb blue light — blue itself. He is now generalizing this photochemistry. DREADCO's 'UniPaint' will contain a broad mixture of dyes embracing the whole spectrum; initially, therefore, it will be black. But cunningly, each dye is bleached by light of all colours but its own. The user will coat the product onto a scratched or damaged area of his car, and quickly cover the wet surface with an opaque patch. Light will leak into this covered region round the edges of the patch, through the original paint. It will therefore be filtered to that exact colour, and will bleach away those dyes in the adjacent UniPaint that absorb that colour. The remaining ones will, of course, give the UniPaint precisely that same shade, which will filter inwards by transmission to colour the UniPaint further in, and so on. Ultimately, the outer colour will infiltrate and transform the entire painted area. The user will peel off the patch to find a dry and perfectly matched repair.

The obvious snag is that such a light-sensitive paint would soon be bleached white by daylight, ruining the match. Daedalus will counter this by coupling the bleaching process to the setting reaction of the paint. While UniPaint is wet, the bleaching agent in its formulation can diffuse to each photo-excited dye molecule and fix it in its colourless state; once it has hardened, no further chemistry is possible and the colour will be stable.

UniPaint will transform, not just the automotive touch-in market, but the whole paint business itself. Its thousands of shades will be replaced by just one. For a start, the car industry could cover every car with UniPaint, and colour it by exposure to the right illumination while the paint was setting. Mixed lighting could even give a graded finish. Indeed, any image at all could be projected onto a wet UniPaint surface, and would be stable when it had dried. Posters, stage scenery, interior decor, street notices, each could be projected from a big slide-projector onto a freshly UniPainted surface during the hours of darkness, to blaze forth in multicoloured splendour in the morning.

David Jones

Geodesic surfactant structures

SIR — Structures formed by the self-association of amphiphiles (liquid crystals, vesicles and micelles) excite interest because of their potential uses as vectors for drug and gene delivery (see, for example, ref. 1), and as matrices from which new structured materials can be made (for example, ref. 2). We have used spherical

nonionic surfactant vesicles (niosomes)³, which are synthetic analogues of liposomes, as vehicles for drug delivery^{4,5}, and have identified in some of them mixtures of nonionic surfactants with cholesterol and either cholesteryl(24)polyoxyethylene ether or dicetyl phosphate. These structures include large (10–50 μm) disk-shaped vesicles, or discoms⁶. In studies on a range of niosome dispersions, freeze-fracture electron micrographs reveal related novel vesicle aggregates, which look like geodesic spheres or raspberries, respectively, and which we term 'multi-niosomes' (see figure).

Niosomes were prepared from sorbitan mono-oleate (Span 80), cholesterol and dicetyl phosphate (in the molar ratio 19:19:2) by dissolving the lipids in chloroform, removing the organic solvent and drying the lipid film with oxygen-free nitrogen. The dried lipid film was hydrated with distilled water and the resultant multilamellar vesicles left to cool to room temperature. The system was quenched rapidly using the sandwich technique, and metal-shadowed at -120°C in a Balzers BAF 400D freeze-fracture system.

Examination of the cleaned replicas in a Jeol 100B electron microscope⁷ revealed previously unreported structures (see figure). The most striking of these (*a* in the figure) is clearly spherical, and its surface is covered with facets, mostly flat, whose straight boundaries are deeply etched. These resemble the framework of a geodesic dome, suggesting that they are formed from a monolayer of close-packed niosomes on the surface of a large (2 μm) spherical vesicle, similar to the 'wafer structure' seen in the membranes of *Streptomyces hygroscopicus*⁸, and thought to result from the hexagonal or cubic organization of membrane phospholipids between the lamellae of artificial lipid membranes.

The 'raspberry' structure (*c* in the figure) is apparently not spherical, perhaps a result of internal packing of the small spherical vesicles, either through hexagonal close-packing or in a face-centred cubic lattice. Near and parallel to the right-hand edge is a row of three squares, each with full-size spheres at the corners and smaller spheres in the centres. The smaller spheres could be the tips of vesicles from the layer beneath, positioned according to a face-centred cubic lattice. The hexagonal packing of full-size spheres over most of the bottom could represent the (111) face of a face-centred cubic lattice.

The best explanation for the raspberry

structures thus seems to be that they are an aggregate of small spheres (niosomes) packed in a somewhat disordered face-centred cubic lattice. It would be simplest if the three structures in the figure were fundamentally similar, that is, that they were all niosomes close-packed to give an aggregate like a solid foam (with the gas replaced by water). The flocculation of small vesicles, either together or with a much larger vesicle, is perhaps caused by something akin to a phase separation at low temperature. Nonionic vesicles containing dicetyl phosphate have a surface negative charge, and the sorbitan head groups will have a low degree of hydration which would otherwise provide some degree of enthalpic repulsion on close approach of the vesicles at room temperature. A better understanding of such vesicle-vesicle interactions may shed light on cell-cell interactions.

B. Sternberg

Friedrich-Schiller University,
Zieglmuhlenweg 1,
Jena D07743, Germany

M. F. Moody

T. Yoshioka*

A. T. Florence

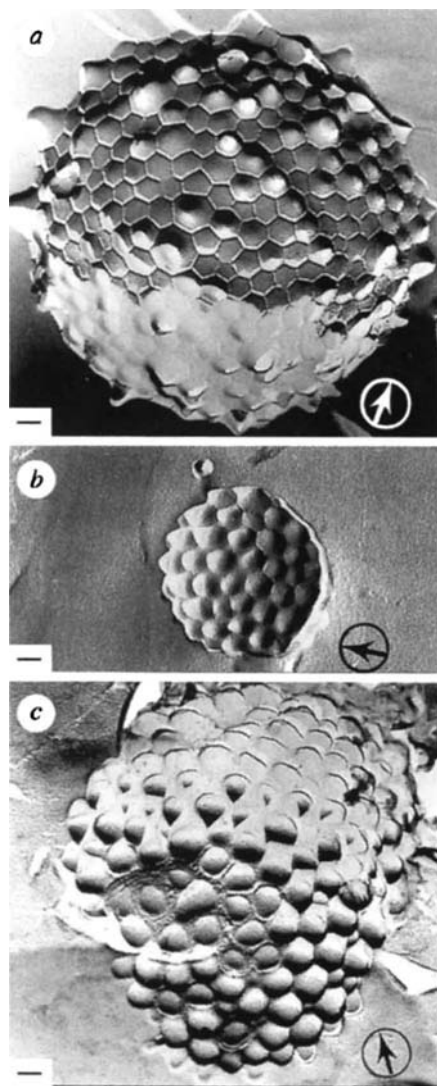
School of Pharmacy,
University of London,
29/39 Brunswick Square,
London WC1N 1AX, UK

1. Caplen, N. J. *et al.* *Nature Med.* **1**, 39–46 (1995).
2. Monnier, A. *et al.* *Science* **261**, 1299–1303 (1993).
3. Baillie, A. J., Florence, A. T., Hume, L. R., Muirhead, G. T. & Rogers, A. J. *Pharm. Pharmac.* **37**, 863–868 (1985).
4. Azmin, M. N. *et al.* *J. Pharm. Pharmac.* **37**, 237–242 (1985).
5. Rogers, A., Cumming, J., Willmott, N. & Florence, A. T. *J. Pharm. Pharmac.* **40**, 337–342 (1988).
6. Uchegbu, I. F., Bouwstra, J. & Florence, A. T. *J. phys. Chem.* **96**, 10548–10553 (1992).
7. Sternberg, B. in *Liposome Technology* Vol. 1 (ed. Gregoriadis, G.) 363–383 (CRC, Boca Raton, 1993).
8. Sternberg, B., Grumpert, J., Reinhardt, G. & Gawrich, K. *Biochim. biophys. Acta* **898**, 223–230 (1987).

HOT and the North Pacific gyre

SIR — Karl *et al.*¹ report changes in conditions at the Hawaiian Ocean Time Series (HOT) station ALOHA ($22^\circ 45' \text{N}$; 158°W) and that these indicate a significant change in the structure and productivity of the pelagic ecosystem of the subtropical North Pacific. The changes were associated with and perhaps due to an intensified near-surface stratification and a shallower mixed layer. This station is, however, unlikely to be "representative of the oligotrophic North Pacific Ocean habitat", as stated by Karl *et al.*

Station ALOHA was established to "assess and interpret annual- to decadal-scale ecosystem variability in the North Pacific subtropical gyre"¹. Measurements there "are designed to obtain oceanographic data on the physics, chemistry and biology of this representative North Pacific



Freeze-fracture electron micrographs (bars represent 100 nm; shadow direction marked by circled arrow). *a*, A geodesic multiniosome. *b*, An intermediate state; the spherical shape is less clear, with no radial foreshortening (except perhaps at the left edge). Facets are convex and their boundaries are mostly unclear, except for one vertical etched line near the middle. *c*, A raspberry structure found in dispersions of sorbitan mono-oleate (Span 80). In the geodesic multiniosome, the vesicles are tightly packed and in the raspberry structure they are more loosely arranged, held together by an envelope of one or several membranes (*c*). In the intermediate state (*b*) the pressure leads to a higher degree of order in the raspberry structure (left side) and shows some features of the geodesic multiniosome form (right side).

*Present address: Snow Brand Milk Products Co Ltd, Technical Research Institute, 1-1-2 Minamidai, Kawagoe, Saitama 350, Japan.

ic habitat." But Karl *et al.* cite no evidence that station ALOHA is "representative" of the oligotrophic North Pacific subtropical gyre, and there is strong evidence against this notion.

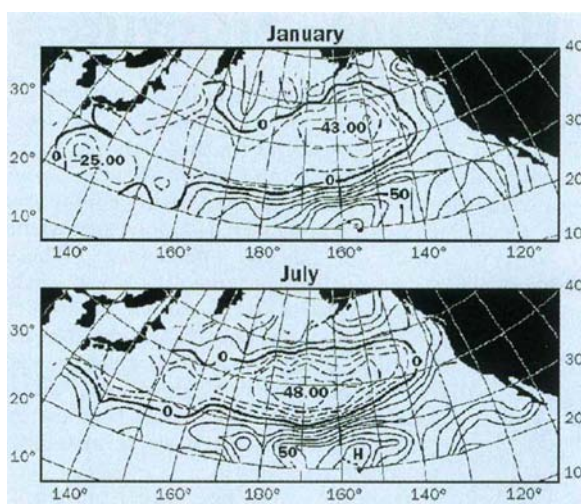
For more than 20 years, persistent, large-scale, coherent sea-surface temperature anomaly patterns have been reported in the North Pacific². These are associated with long-lasting atmospheric anomaly patterns. Large sectors of positive anomaly exist concurrently with other large sectors of negative anomaly. Most of the central anticyclonic gyre is out of phase with eastern boundary currents and the ocean south of about 20° N (ref. 3). The HOT ALOHA station lies in the steep gradient between these two regimes (see figure). Monthly sea-surface temperature anomalies averaged over the 5°×5° surrounding station ALOHA are not correlated, either positively or negatively, with 5°×5° temperature anomalies over most of the central, oligotrophic gyre. These large-scale temperature anomalies at the sea surface are known to extend to at least the upper pycnocline and are accompanied by changes in upper water column physical and biological structure⁴⁻⁷.

If ecosystem properties such as productivity, nutrient cycling and biomass vary in response to climatic and water column changes of the magnitude indexed by the sea-surface temperature anomaly patterns, then it is evident that studies at station ALOHA are not representative of the anticyclonic central gyre.

John McGowan

University of California,
San Diego Marine Life Research
Group, 0227,
Scripps Institution of Oceanography,
La Jolla, California 92093-0227, USA

KARL REPLIES — The ocean is variable on nearly all time and space scales that have been examined to date⁸. Consequently, the design of oceanographic field studies must accommodate this anticipated variability to avoid introducing a sampling or data interpretation bias. Ideally, a sufficient number of 'representative' stations in a given habitat must be revisited frequently over at least a decade, or longer. Unfortunately, this ideal sampling strategy is not practical, so intelligent com-



Contours of seasonal cross-correlations of sea-surface temperature anomaly of the 5°×5° surrounding the HOT ALOHA station (22°45' N; 158° W) with all other 5°×5° anomalies in the North Pacific during 1947–94. Solid lines are positive, broken lines negative. H indicates the 5°×5° that includes station ALOHA (data provided by D. R. Cayan).

promises must be made. One approach is to establish time-series sampling stations at strategic sites that are characteristic of key oceanic habitats. This strategy was adopted by the US Joint Global Ocean Flux Study (JGOFS) programme and led to the establishment of oligotrophic ocean benchmarks near Bermuda⁹ and Hawaii¹⁰.

To date, this combined JGOFS time-series research effort of >150 research cruises on approximately monthly intervals has provided an unprecedented view of the complexity of physical and biogeochemical processes in these distinct subtropical habitats. Of general importance are the discoveries of relatively high annual rates of primary production (around 14 mol C m⁻² yr⁻¹), the importance

of N₂ fixation as a source of 'new' nitrogen and the temporal decoupling of organic matter production, export and decomposition. A hallmark of these oligotrophic habitats is variability, a remarkable departure from the previous view of a 'monotonous self-regulating climax ecosystem'⁵ based on research conducted near 28° N, 155° W. Of the 18 cruises during 1971–85 (ref. 11), approximately 70% were in summer (June–September) and 5% in August alone. No cruises were conducted in 1970, 1975, 1978–79, 1981 or 1984, further time-biasing this data set.

Station ALOHA, the HOT programme site, is located at 22°45' N, 158° W in deep water (4,750 m),

one Rossby radius (50 km) upwind of steep topography associated with the Hawaiian Ridge, but close enough to Honolulu to provide feasible logistical support. Based on an analysis of sea-surface temperature anomaly patterns over the North Pacific Ocean, McGowan disputes that our results from station ALOHA are truly representative of the anticyclonic central gyre. While this opinion may be valid, without the presentation of an equivalent, synoptic oceanographic data set on biological rates and processes from a 'representative' station, it cannot be proved definitively.

David M. Karl

Department of Oceanography,
University of Hawaii,
Honolulu, Hawaii 96822, USA

Evolution of deuterium on Venus

SIR — Venus has a large deuterium-to-hydrogen (D/H) ratio, almost 160 times that of Earth¹. Current models assume that Venus initially had an Earth-like D/H ratio which has risen due to atmospheric evolution. Two models have been advanced: the primordial-ocean model, which interprets the high D/H ratio as the result of the fractionation loss from a large initial reservoir of water (perhaps even an 'ocean'), and the steady-state model, where the ratio results from a balance between sources of water (probably volcanic) and fractionation losses, with the total abundance of water remaining constant over time². The primordial-ocean model allows extrapolation back to the initial water abundance, whereas the steady-state model implies that this information has been lost.

Grinspoon^{2,3} has argued that the steady-state model provides a good

description of the current water budget, as the timescale for D/H evolution is more favourable to a steady-state solution. The steady-state model predicts a limiting D/H ratio of α/f , where α is the source D/H ratio (from, for example, outgassing) and f is the fractionation factor (describing the efficiency of D relative to H escape; f is currently 0.13; ref. 4). This creates a paradox of sorts, if the source D/H ratio is terrestrial-like, then the Venus D/H ratio should be only a factor of eight times that on Earth.

Grinspoon has proposed two solutions: either the source D/H is high, enriched by

- Karl, D. M. *et al.* *Nature* **373**, 230–234 (1995).
- Namias, J. & Born, R. M. *J. geophys. Res.* **75**, 5952–5955 (1970).
- Namias, J. *et al.* *J. Climate* **1**, 682–703 (1988).
- McGowan, J. A. & Hayward, T. L. *Deep-Sea Res.* **25**, 771–793 (1978).
- McGowan, J. A. & Walker, P. W. *Ecol. Monogr.* **49**, 195–226 (1979).
- Vernick, E. L. *et al.* *Science* **238**, 70–72 (1987).
- Roemmich, D. *Science* **257**, 373 (1992).
- Dickey, T. *Rev. Geophys.* **29**, 383–413 (1991).
- Michaels, A. F. *et al.* *Deep-Sea Res.* **41**, 1013–1038 (1994).
- Karl, D. M. & Winn, C. D. *Envir. Sci. Technol.* **25**, 1976–1981 (1991).
- Hayward, T. L. *Deep-Sea Res.* **34**, 1593–1627 (1987).

- Donahue, T. M. & Hodges, R. R. *Geophys. Res. Lett.* **20**, 591–594 (1993).
- Grinspoon, D. H. *Science* **238**, 1702–1704 (1987).
- Grinspoon, D. H. *Nature* **363**, 428–430 (1993).
- Donahue, T. M. & Hartle, R. E. *Geophys. Res. Lett.* **19**, 2449–2452 (1992).

an early, massive fractionation loss of water, the residual being locked within the mantle, slowly outgassing; or a recent ($<10^9$ yr ago) catastrophic resurfacing of Venus released a large amount of water which was highly fractionated by atmospheric loss processes, resulting in a transitory, enriched D/H that is slowly returning to the steady-state value.

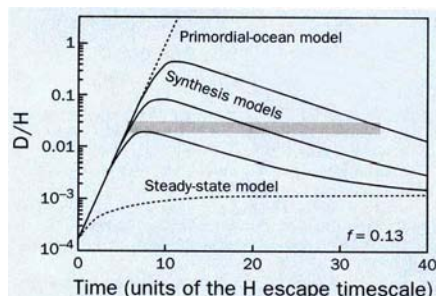
An alternative solution can be achieved from a synthesis of the primordial-ocean and steady-state D/H models. In this, it is assumed that water is in a steady state now (or in the future), but that a steady state has not always dominated. The synthesis D/H evolution model can be described by:

$$(D/H)t = (\chi Re^{-f} + (\alpha/f)(1-e^{-f})) / (Re^{-f} + (1-e^{-f}))$$

where χ is the initial D/H ratio, R is the ratio of the initial water abundance to the eventual steady-state water abundance, and time is given in units of the hydrogen escape timescale, which is between 125 and 800 Myr (ref. 4; derivation available on request). This is a more general solution than the previous models, but reduces to the primordial-ocean model for $Re^{-f} \gg (1-e^{-f})$ and to the steady-state model for $R = 1$.

The figure shows several time evolutions for D/H, assuming that the initial and source D/H ratios are Earth-like. The simple steady-state model cannot intercept the current Venus atmosphere (represented by the bar in the figure) unless the initial and/or source D/H ratio is much larger than terrestrial, as Grinspoon noted. The primordial-ocean model does intercept the current state if the hydrogen escape timescale is near its maximum, but if it is shorter the primordial-ocean model predicts a much higher D/H ratio than is observed.

The synthesis model has characteristics of both models, showing strong early fractionation (similar to the primordial-ocean model), and then asymptotically



Evolution of D/H for the primordial-ocean, steady-state and synthesis models. The bar represents the possible range for current Venus atmosphere age and D/H ratio. The large uncertainty in the H escape timescale causes the large range in the age of Venus. Three curves generated by the synthesis model, with R equal to 14,000 (upper curve), 2,500 (middle curve) and 360 (lower curve), are shown.

approaching a steady-state value (as in the steady-state model). The critical transition and the maximum D/H ratio occur when the hydrogen abundance has evolved to within a factor of around 3 of the eventual steady-state abundance.

The synthesis model intercepts the current Venus state for R from 360 to 14,000. For R less than 360, the current D/H ratio is unreachable, given an initial D/H ratio that is Earth-like.

The model emphasizes that a minimum ratio of the initial to eventual steady-state water abundance of 360 is required to

match the current D/H ratio. This implies that the initial water abundance was a minimum of 120 times the current water abundance, or the equivalent of a 1.8-m global ocean. These are conservative estimates assuming constant f ; f was probably higher in the past, allowing even higher values of R , perhaps as much as that for a terrestrial ocean.

Mark A. Gurwell

Harvard-Smithsonian Center for

Astrophysics,

60 Garden Street,

Cambridge, Massachusetts 02138, USA

Ice cores and mammoth extinction

SIR — The detailed records of past climate revealed by ice cores and other recent studies¹⁻⁴ have important implications for our understanding of the causes of the extinction of mammoths and other large mammal species around the Pleistocene/Holocene boundary.

The Late Pleistocene Holarctic 'mammoth fauna' consisted of a unique combination of grazing species intimately linked to a peculiar tundra-steppe environment, characterized by a dry, cold, continental climate and a complex, productive mosaic vegetation dominated by grassland. This ecosystem has no modern analogue, and many workers have suggested that its decline, and replacement by the relatively low-diversity, less productive Holocene zonation of tundra and boreal forest, led to the extinction of many megafaunal species^{5,6}.

It has been argued, on circumstantial evidence, that the higher biotic diversity of the Pleistocene might in some way reflect greater climatic variability than in the Holocene⁵. Recent discoveries seem to support this model. The Greenland ice cores indicate almost constant and rapid climatic fluctuations over the 100 kyr before the Holocene warming and, in strong contrast, a stable climatic signal for the Holocene itself¹. Conceivably, the constant fluctuations of climate in the Late Pleistocene were partly responsible for maintaining a mosaic of plant communities, by a constant 'stirring', favouring a pioneering character of vegetation which supported the grazing megafauna⁷. In contrast, the unique stability of Holocene climate, indicated by the ice cores, may have contributed to the development of today's strong zonation of climax vegetation types, such as tundra and coniferous forest, unsuitable for large grazers like the woolly mammoth.

The detailed record of mammoth extinction is in general agreement with the climatic changes seen in the ice cores. Before 14 kyr (12.5 radiocarbon kyr), the woolly mammoth inhabited the whole of northern Eurasia. At around that time, in keeping with broad-scale warming and

forest spread in northern Eurasia, the mammoth's range became restricted to Arctic Siberia, as if 'retreating' to the north⁸. Although the Arctic also experienced important vegetational change, mammoths could survive there for a few thousand years. Around 11.5 kyr, however, there was a second major event, evidenced by the appearance of a myriad of thaw lakes all over the unglaciated Arctic⁸, marking the final demise of the tundra-steppe biome.

This event corresponds to the beginning of the period of stability documented by the curve from the GRIP ice core from Summit, Greenland, as a rise at 11.5 kyr, reaching a plateau at around 10 kyr. By 10.5 kyr (9.0 radiocarbon kyr), mammoths were extinct everywhere except the high Arctic island of Wrangel, where a dwarfed, insular population survived on relict tundra-steppe vegetation⁹.

One problem of the climatic model of extinction has been to explain how woolly mammoth and other tundra-steppe mammals survived earlier interglacials, generally interpreted as climatic cycles of afforestation analogous to the Holocene. Some palaeontologists have proposed that the vegetation of the interglacials, in fact, differed from that of the Holocene. In northeast Siberia, for example, several 'forest' intervals correlated with interglacials have been recorded, but their vegetation differed from modern larch-dominated taiga. The main tree species was birch, and there was a substantial grass component, while insect faunas preserved some steppe

1. Dansgaard, W. *et al.* *Nature* **364**, 218-220 (1993).
2. Thouveny, N. *et al.* *Nature* **371**, 503-506 (1994).
3. Field, M. H., Huntley, B. & Müller, H. *Nature* **371**, 779-783 (1994).
4. Cortijo, E. *et al.* *Nature* **372**, 446-449 (1994).
5. Guthrie, R. D. in *Quaternary Extinctions: A Prehistoric Revolution* (eds Martin, P. S. & Klein, R. G.) 259-298 (Univ. Arizona, Tucson, 1984).
6. Sher, A. V. *Zh. Obshch. Biol.* **51**, 163-177 (1990).
7. Lister, A. M. & Bahn, P. G. *Mammoths* (Boxtree, London, 1995).
8. Sher, A. V. in *The 22nd Arctic Workshop 1992*, 125-127 (Univ. Colorado, Boulder, 1992).
9. Vartanyan, S. L., Garutt, V. E. & Sher, A. V. *Nature* **362**, 337-340 (1993).
10. Sher, A. V. *Q. J. Intern.* **10/12**, 215-222 (1991).
11. Zahn, R. *Nature* **371**, 289-290 (1994).

species, indicating a habitat that could support grazing mammals¹⁰.

The GRIP core seemed to corroborate a difference between Holocene climate and that of previous interglacials by showing major fluctuations within the 'last interglacial' around 120 kyr. The veracity of this part of the core has been questioned, however: the 'fluctuations' may be due to folding; moreover, they are not seen in some other ice-core and oceanic sequences¹¹. On the other hand, independent lines of evidence corroborate Eemian inconstancy, including data from the crucial terrestrial realm. Data from a lake core in France² indicate two rapid cooling events within the last interglacial, including a period of reversion from forest to open subalpine vegetation. Analysis of detailed pollen records from Germany³ suggests greater climatic instability in the last interglacial than in the Holocene. Finally, evidence of mid-Eemian cooling has come from studies in the Norwegian Sea and North Atlantic Ocean⁴.

The spatial structure of the terrestrial interglacial environment requires further investigation, as does the mammoths' response to it, which clearly varied across their vast range. A crucial role may have been played by unglaciated Arctic regions in providing interglacial refuges for the tundra-steppe grazers. This pattern was again followed at the start of the Pleistocene/Holocene transition, which seems to have begun as a 'normal' interglacial. However, subsequent changes, culminating in the onset of stable climate documented by the ice cores, resulted in an irreversible turnover of the Pleistocene biota.

Adrian M. Lister

*Department of Biology,
University College London,
London WC1E 6BT, UK*

Andrei V. Sher

*Severtsov Institute of Ecology
and Evolution,
Russian Academy of Sciences,
Moscow 117071, Russia*

original excavation, found at the museum.

We used mass spectrometric U-series analysis of the enamel and dentine to evaluate U-uptake history, which affects the calculated ESR age⁵. Early uptake yields the minimum age for a given data set, while linear-uptake ages commonly agree with other age estimates. Equivalent early-uptake ages and U-series ages in teeth from layers 5–6 indicated early U-uptake, whereas early-uptake ages of teeth from layers 7–8 were larger than associated U-series dates, implying continuous U uptake. The respective linear-uptake or early-uptake ages of teeth from layers 1–8 were indistinguishable, with a mean of 130 ± 10 kyr, showing that the entire fluvial sequence was formed in the last interglacial (stage 5e), consistent with its faunal character.

Whereas hominids from beds 3–7 are all Neanderthals, layer 8 (also dated to 130 kyr) contained cranial elements of a 6–8-yr-old juvenile (Krapina 1). Some metric and morphological aspects group it with other juvenile Neanderthals, but its more pentagonal shape in rear view, higher frontal squama and some other characteristics approach the condition of early modern humans (N. Minugh-Purvis, unpublished thesis, Univ. Pennsylvania). A partial occipital bone from layer 8 (Krapina 11) lacks a suprainiac fossa, a characteristic of Neanderthals observed in occipitals from layers 3–4. These data suggest that the layer 8 hominids are possibly transitional to anatomically modern hominids, raising the question of when this taxon first appeared in Europe.

W. J. Rink

H. P. Schwarcz

*Department of Geology,
McMaster University, Hamilton,
Ontario L8S 4M1, Canada*

F. H. Smith

*Department of Anthropology,
Northern Illinois University,
DeKalb, Illinois 60115, USA*

J. Radović

*Croatian Natural History Museum,
Demetrova 1, 1000 Zagreb, Croatia*

1. Smith, F. in *The Origins of Modern Humans* (eds Smith, F. H. & Spencer, F.) 137–209 (Liss, New York, 1984).
2. Gorjanović-Kramberger, D. *Der diluviale Mensch von Krapina in Kroatien: Ein Beitrag zur Paläoanthropologie* (Krieger, Wiesbaden, 1906).
3. Simek, J. in *Raw Material Economies Among Prehistoric Hunter Gatherers* (eds Montet-White, A. & Holen, S.) 59–71 (Univ. Kansas Publ. Anthropology, Lawrence, KS, 1991).
4. Grün, R., Schwarcz, H. P. & Zymela, S. *Can. J. Earth Sci.* **24**, 1022–1037 (1987).
5. Grün, R., Schwarcz, H. P. & Chadam, J. J. *nucl. Tracks Radiat. Measmt* **14**, 237–241 (1988).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and five references.

ESR ages for Krapina hominids

SIR — At the site of Krapina, 40 km NNW of Zagreb, more than 850 skeletal remains from several Neanderthal individuals were found, principally in layers 3 and 4 (refs 1, 2), associated with a Middle Palaeolithic industry³. We have obtained electron spin resonance (ESR)⁴ and U-series dates on tooth enamel from

the site (see table). Teeth from the 1899–1905 archaeological excavations of K. Gorjanović-Kramberger, stored at the Croatian Natural History Museum, were analysed as in ref. 4. External γ - and β -dose rates were estimated from U, Th and K analyses of blocks of weakly indurated sand preserved from the

ESR AND U-SERIES DATES FOR KRAPINA

Layer	Sample	ESR age (kyr)			U-series age (kyr)		
		Early uptake	Average	Linear uptake	Average	Enamel	Dentine
9	91187A	87±7	87±7	113±10	113±10		
8	91182A	109±9		147±14			
	91182B	90±8		120±12			
	91182C	88±7		122±11			
	91182D	90±6		126±11			
	91182E	94±6		128±11			
	91182G	88±8		120±11			
	91182H	93±9		130±12			
	91182I	90±10		123±13			
	91183A	108±9	99±5	139±14	132±5	69.7±0.5	89.8±1.4
7–8	91183B	96±8		128±12			
	91183C	94±7		127±12			
	91183D	94±7		128±12			
	91183E	95±8		126±12			
	91183F	103±9		135±13			
	91183G	100±8		132±13			
	91183H	102±8		135±13			
	91184	131±13		152±17			
	91184	131±13		152±17			
5–6	91186A	126±11	130±3	178±17	167±10	131±1	223±4 ⁺⁵⁵
	93042A	132±13		160±18			415 ^{±38}
	93042B	132±14		162±19			
1	91185A	142±12	133±15	185±19	180±15		
	91188A	142±11		192±18			
	91189A	116±9		163±16			

ESR dates are computed assuming $10 \pm 10\%$ water in sediment. γ - And β -doses from sediment were determined from neutron activation analyses of sediment samples for U, Th and K. β -doses were corrected for attenuation. U-series ages were determined by thermal ionization mass spectrometry⁵.

Seismic evidence for a fossil mantle plume beneath South America and implications for plate driving forces

John C. VanDecar^{*}, David E. James^{*} & Marcelo Assumpção[†]

^{*} Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington DC 20015, USA

[†] Instituto Astronômico e Geofísica, Departamento de Geofísica, Universidade de São Paulo, São Paulo, Brazil

A teleseismic travel-time study reveals the presence of a fossil plume in the deep upper mantle beneath Brazil, which has apparently remained geographically fixed with respect to the overlying continent despite thousands of kilometres of plate motion. This result implies that the upper mantle and lithosphere beneath South America have remained coupled since the breakup of Gondwanaland and may provide an answer to the long-standing question of what forces drive continental plates.

ALTHOUGH the theory of plate tectonics has been successful at explaining many of the major geological features observed on the Earth's surface, the relationship of plate motions to dynamical processes in the sub-lithospheric mantle remains controversial. In classical plate-tectonic theory, the motion of lithospheric plates is considered to be essentially independent of flow in the upper mantle beneath¹. But it has been suggested² that some plate motions may be coupled to, or even driven by, upper-mantle flow. Such large-scale flow may be particularly important for large continental plates, where driving mechanisms such as slab pull or ridge push appear inadequate to move the plate, and where entrainment of deep lithospheric keels³ in upper-mantle flow may be an important factor in plate motion.

This study is part of the Brazilian Lithospheric Seismic Project⁴, undertaken in 1992 with the installation of ten portable matched three-component broadband digital seismograph systems across the major Precambrian provinces of southeastern Brazil. The instruments were deployed in an approximately east-west band along 20–30° S latitude, with an east-west aperture of approximately 800 km (Fig. 1). The purpose of the experiment was to investigate three-dimensional lateral variations in deep lithospheric structure beneath four highly heterogeneous continental segments of the Brazilian shield—the Archaean São Francisco craton, the late Proterozoic Brasília mobile belt, the intracratonic Paraná basin, and the coastal Ribeira shear belt⁵.

In planning the Brazil experiment, we did not anticipate observing large-scale heterogeneity in the mantle beneath the Brazilian lithosphere, for stable shield areas are generally thought to exhibit relatively little lateral variation in the sub-lithospheric mantle. It was therefore a most unexpected result when our first travel-time inversions for three-dimensional P- and S-wave velocity structure revealed that the strongest and most robust anomaly in the region is a large-scale low-velocity structure shaped roughly like a vertical cylinder in the upper mantle below about 200 km. The fact that this remarkable plume-like feature extends to at least 500–600 km depth and that it lies beneath the great Paraná flood basalt province (age ~130 Myr) leads us to suggest that it represents a 'fossil' expression of the plume conduit in the deep upper mantle that supplied the Paraná plume head.

The existence of a fossil plume conduit in this region is wholly unpredicted in that South America has a high absolute plate velocity (3.5 cm yr⁻¹)⁶, so that all residual thermal and chemical evidence for the plume in the deep (sub-lithospheric) upper mantle should have been left several thousand kilometres behind the moving plate over the course of 130 Myr. Thus the persist-

ence of a plume trace in close geographical proximity to the original position of the ascending plume head beneath the Paraná basin implies that the mantle, to a depth of at least 500 km, has been in coupled motion with the overlying South American plate since the opening of the South Atlantic Ocean in early Cretaceous time. Our seismic results lend support to recent speculations⁷ that the South American plate could be coupled to sub-lithospheric upper-mantle flow, and may therefore help us to understand the nature of the forces that drive continental plates.

Plume history

Between about 137 and 80 Myr ago, the region of study and areas to the south underwent a series of plume-related volcanic episodes. Activity commenced with a massive outpouring of continental flood basalts in the Paraná basin over the time interval 137–127 Myr, just before the separation of South America from Africa (~120 Myr)^{8–10}. The flood basalts are widely supposed to be related to the ascending head of the Tristan da Cunha plume. By about 126 Myr ago¹⁰, the plume tail had migrated to the Mid-Atlantic Ridge to form the Tristan da Cunha hotspot. Continued hotspot activity at Tristan da Cunha produced both the Rio Grande rise and the Walvis ridge¹¹. The Paraná basin flood basalts were followed in late Cretaceous time (~80–90 Myr) by alkalic volcanism that was widespread around the northern and eastern margins of the Paraná basin¹² (principal volcanic centres shown as yellow circles in Fig. 1). The origin of these alkalic volcanic rocks, some of which contain mantle xenoliths from depths of 150 km or greater¹³, is controversial and has been attributed by some to a plume head 80–90 Myr in age associated with the Trindade hotspot trace^{14,15}. It is our view, however, that on the basis of the South American plate trajectory¹¹ the Trindade hotspot trace is implausibly far north to have produced the observed alkalic volcanism. Moreover, there are no flood basalts associated with the hypothesized Trindade plume head, and the offshore Trindade volcanics are thought by some to be not a hotspot trace at all, but rather associated with a 'leaky transform fault' (G. Karner, personal communication). We instead interpret the alkalic volcanism as an end-stage of Paraná volcanism, which in the coastal areas may have continued to as recently as 55 Myr. We consider it significant that the locus of this late Cretaceous alkalic volcanism occurs within the weak mobile belts on the boundary of the Paraná basin, in a roughly semicircular pattern around the axis of the seismically observed fossil plume conduit. We suggest that the late Cretaceous alkalic volcanism is due to conductive heat

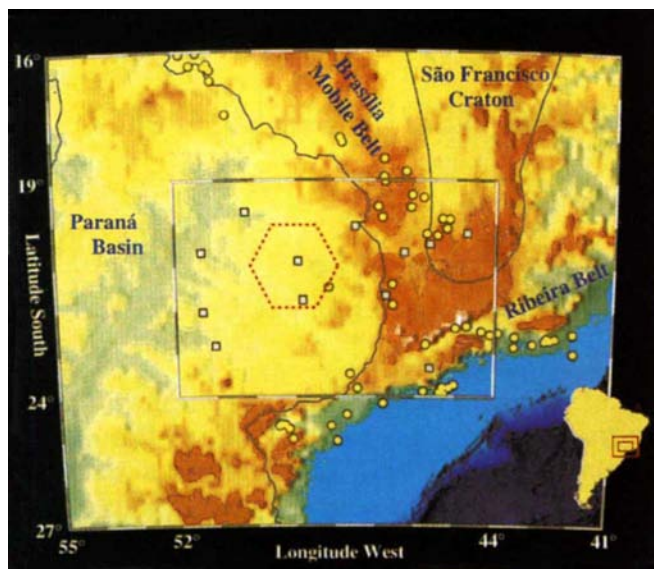


FIG. 1 Location map of the Brazilian broadband seismic stations (white squares) with colour-shaded topography and schematic outline of major geological provinces (solid lines). The exterior boundary defines the region over which the travel-time inversion was performed, while the interior boundary represents the region where appreciable resolution was expected and to which our analysis was confined. The yellow circles represent sites of Cretaceous alkalic volcanism, and the red (dashed) hexagon near the centre delineates the approximate position of the observed deep upper-mantle low-velocity conduit. Inset, location of the target area with respect to the rest of South America.

transport into the lower lithosphere where small degrees of partial melting generated the alkalic magmas. The cessation of alkalic volcanism in the region is not necessarily inconsistent with the persistence of a deep plume signature, as the relative amplitude of the thermal anomaly decays to about 75% its original value over 80 Myr. As we discuss below, the thermally induced density anomaly may, in addition, be balanced by chemical gradients, such as iron enrichment, to produce a neutrally buoyant column that no longer delivers heat to the deep lithosphere by advection.

Obtaining Earth structure from seismic waves

The data used in this study are relative arrival times of compressional (P) and shear (S) waves recorded at 12 sites of the Brazilian broadband network (two of the ten stations were relocated during the experiment to increase station coverage). Absolute times are known to 2 ms (GPS timing). Relative phase arrival times were retrieved via a multichannel cross-correlation procedure that makes use of the duplicate information obtained

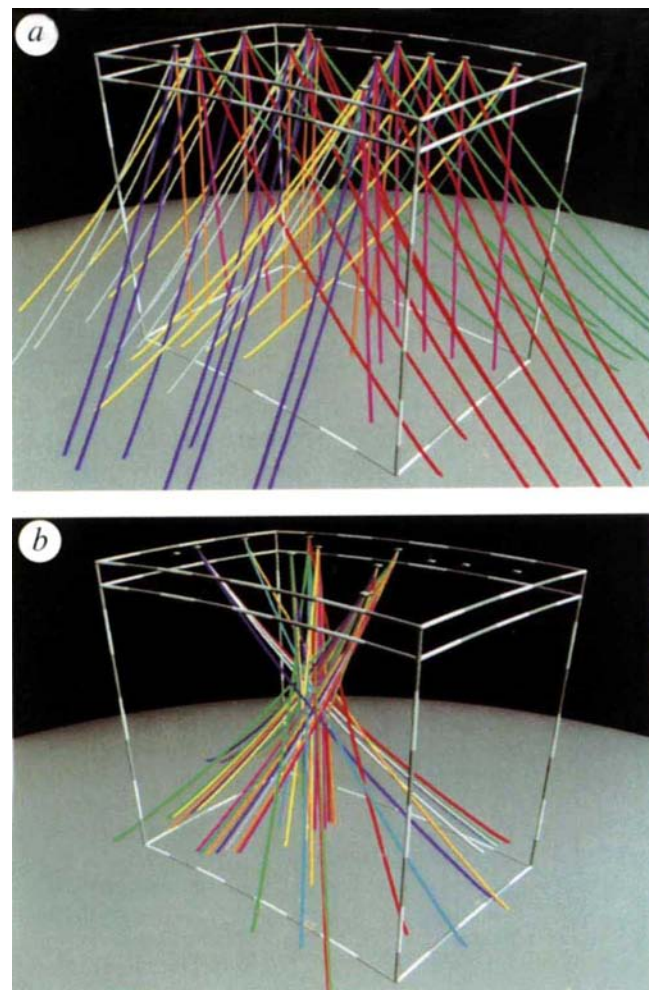
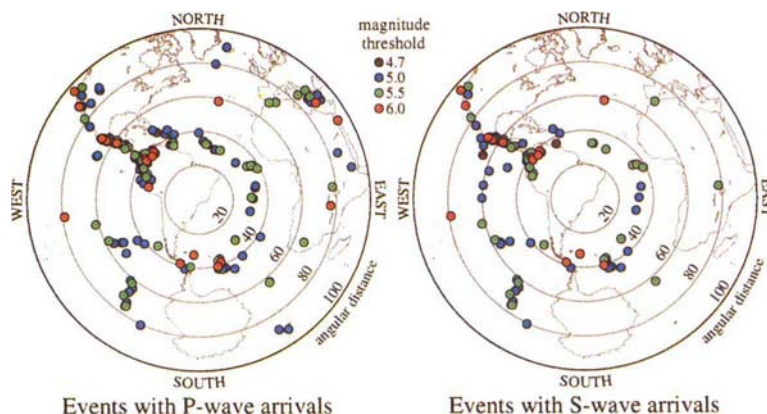


FIG. 3 Ray paths plotted through the upper mantle. The box extends from 0 to 660 km depth (indicated by the underlying grey sphere). *a*, Rays from 8 of the 240 events used in the P-wave inversion (each event is plotted with a different colour); *b*, rays that sampled a small region at 250 km depth at the position of the low-velocity cylindrical anomaly, interpreted as a fossil plume head conduit. The high angles at which ray paths cross at depths between 100 and 500 km enable us to isolate structures located in these regions.

FIG. 2 Location of earthquakes (with respect to the Brazilian seismic network) that resulted in rays which had turning points in the Earth's lower mantle (30–98° angular distance). Shown are the 182 earthquakes from which P-wave observations were used (left) and 113 earthquakes from which S-wave observations were used (right). Four to nine observations were made for each event. We have also used observations from earthquakes whose rays had turning points in the Earth's core (not shown).



by cross-correlating all possible pairs of waveforms¹⁶. This procedure produces both highly accurate delay times (standard errors of ~0.03 s for P waves and 0.10 s for S waves) and useful standard error estimates. The key to the successful use of limited seismic data sets is in obtaining accurate, robust parameters and

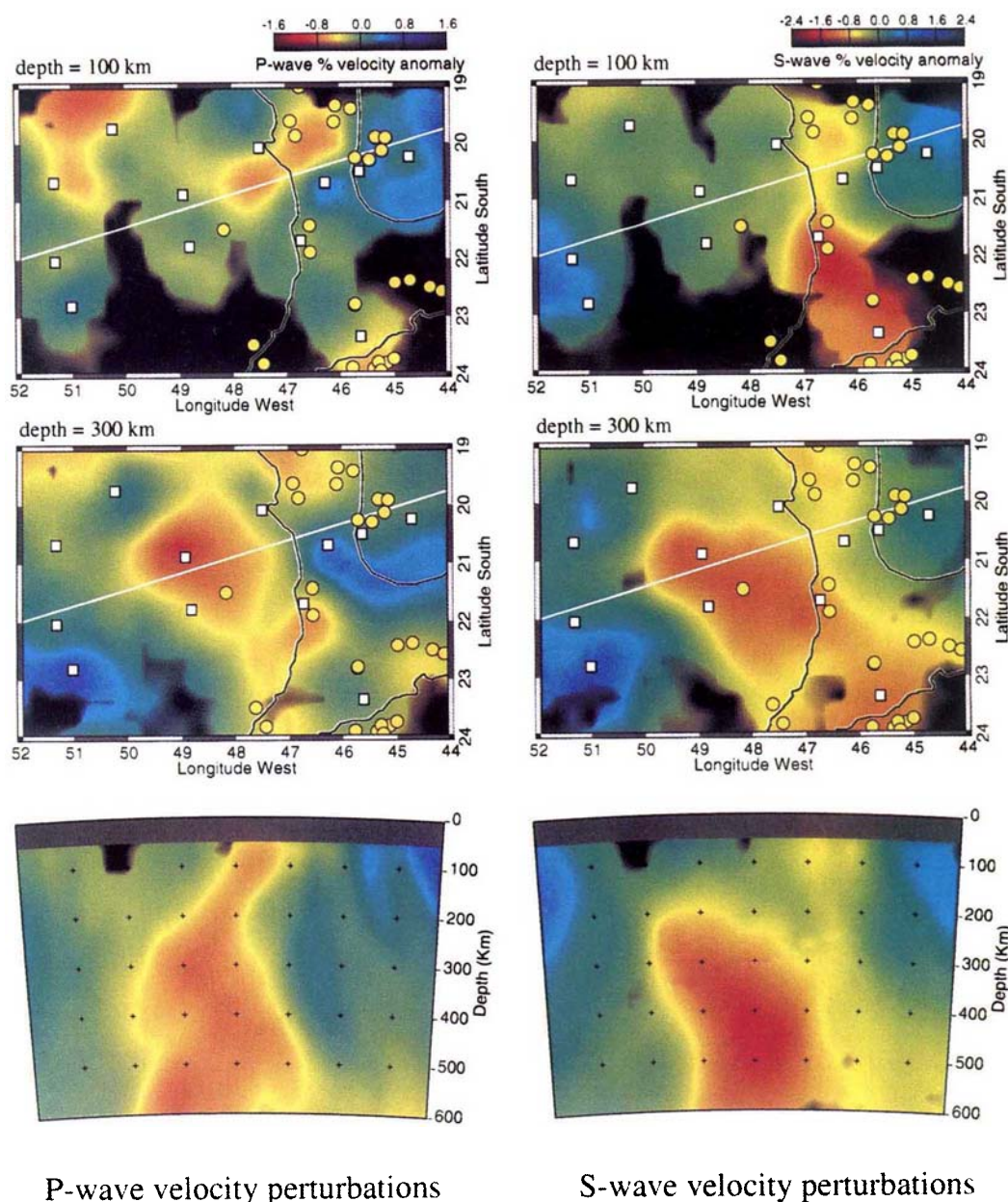


FIG. 4 Cross-sections through the P-wave (left-hand column) and S-wave (right-hand column) velocity perturbation models at constant depths of 100 km (top row) and 300 km (middle row) along with vertical cross-sections (bottom row) at locations shown by white lines in the depth sections. The yellow–red colours represent low-velocity anomalies, and the blue colours represent high-velocity anomalies. The scales fade to black as resolution degrades, with pure black indicating that

no rays traversed those regions. Note the low-velocity, roughly cylindrical, feature rising from deep in the upper mantle, and the high velocities beneath the São Francisco craton (northeastern corner). Also note that the scales are different for the P- and S-wave models— $\pm 1.6\%$ versus $\pm 2.4\%$ —indicating that the S-wave perturbations are in general stronger. The perturbations are with respect to the global radial-Earth velocity model IASP91²⁵. Other symbols are the same as in Fig. 1.

therefore an emphasis of this study was on quality control. Each event was individually analysed with various filter and cross-correlation settings to test for consistency, and to ensure against cycle skipping—always a danger with band-limited data—and biases induced by waveform distortion. In cases where multiple phases arrive closely spaced in time (for example, the S and core-reflected ScS phases) WKB synthetic seismograms¹⁷ were generated and processed in parallel with data to guard against biases being introduced due to phase interference.

Several assumptions are implicit in the inversion analysis presented here. The predominant assumption is that, although the waves that we record have a limited bandwidth, we can make use of the infinite-frequency approximation that allows us to assume that the energy travels from earthquake to seismometer

solely along a ray path—analogue to optical ray theory¹⁷. This approximation is useful when material properties are varying slowly with respect to the wavelength of the seismic wave. The seismic wavelengths in this study are about 8 km for the P waves and 35 km for S waves, whereas the smallest structures that we image have wavelengths of around 100 km. A second assumption is that the problem may be linearized, that is, that our initial guess at earthquake locations and ray paths are linearly close to the real ones or that by iteratively performing linear inversion we will reach this point. To minimize these nonlinear effects, we have chosen to use only earthquakes at teleseismic distances (angular distances greater than 30° from the receivers)¹⁸ so that rays we use are not turning in the steep gradients of Earth's transition zone above 660 km depth.

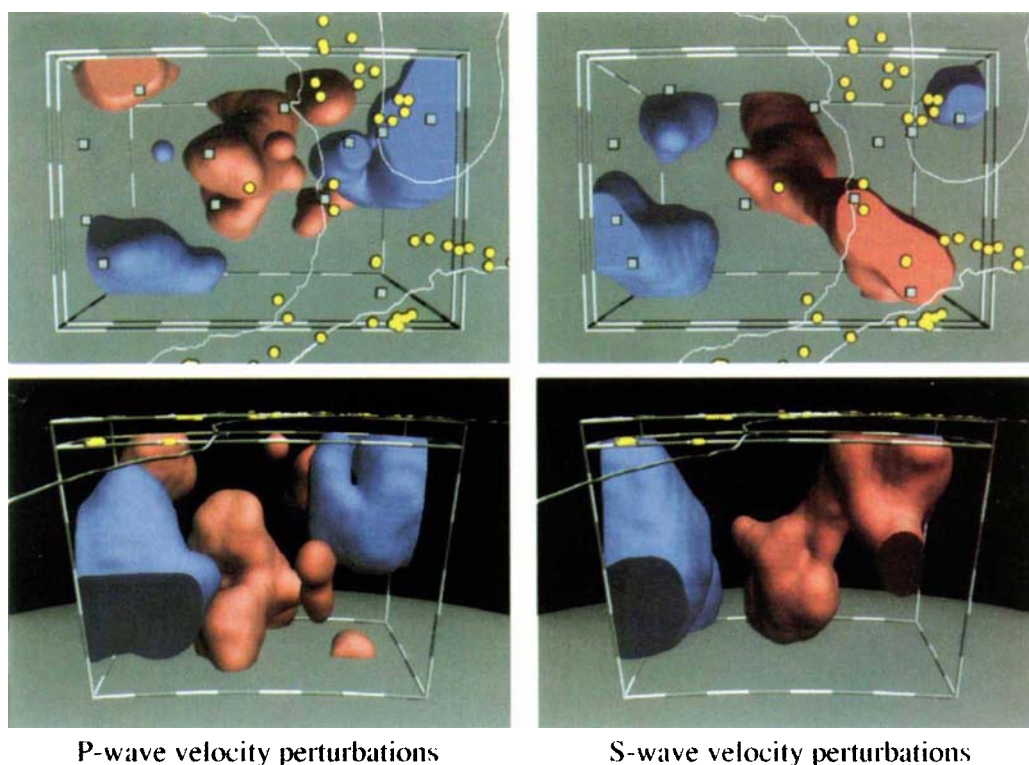


FIG. 5 Perspective views of volumes enclosing high and low P-wave (left-hand column) and S-wave (right-hand column) velocity anomalies. The views are from directly above the target region (top row) and from due South (bottom row). The low-velocity regions are enclosed by red surfaces, and the high-velocity regions by blue surfaces. Each encloses an equal volume of the target region resulting in surfaces at 0.35% high- and 0.65% low-velocity for the P-wave model, and 0.57%

high- and 1.2% low-velocity perturbation for the S-wave model. The velocity model is plotted from 50 to 600 km depth, and in the same lateral dimensions as in Fig. 4. The alternating grey stripes on the boxes indicate 100-km depth marks and degrees in longitude and latitude. The underlying grey sphere is at 660 km depth, and the other symbols are the same as in Fig. 1.

We parametrize variations in Earth structure beneath our stations with splines under tension constrained at a series of regular knots¹⁹. This formulation allows us to have both smooth variations—advantageous for ray-tracing through the resulting models—yet relatively local structure (compared to cubic splines which have large side lobes). We use 32 knots in depth, 36 in latitude and 45 in longitude, giving a total of 51,840. The grid extends from 0 to 1,000 km in depth, 16° to 27° S in latitude, and 41° to 55° W in longitude. The surface extent of this grid is shown by the outside boundary in Fig. 1. The knots are spaced $33\frac{1}{3}$ km apart in depth and one-third of a degree apart in latitude and longitude, allowing us to model structure with a spatial wavelength of about 66 km. It is within the interior portion of the model shown by the internal boundary in Fig. 1 (0–600 km depth, 19–24° S, 44–52° W) that we expect to obtain appreciable resolution and to which we will confine our interpretations. Our model extends to the exterior region to mitigate the mapping of unwarranted and spurious structure into our interior model due to that in outside regions^{19,20}.

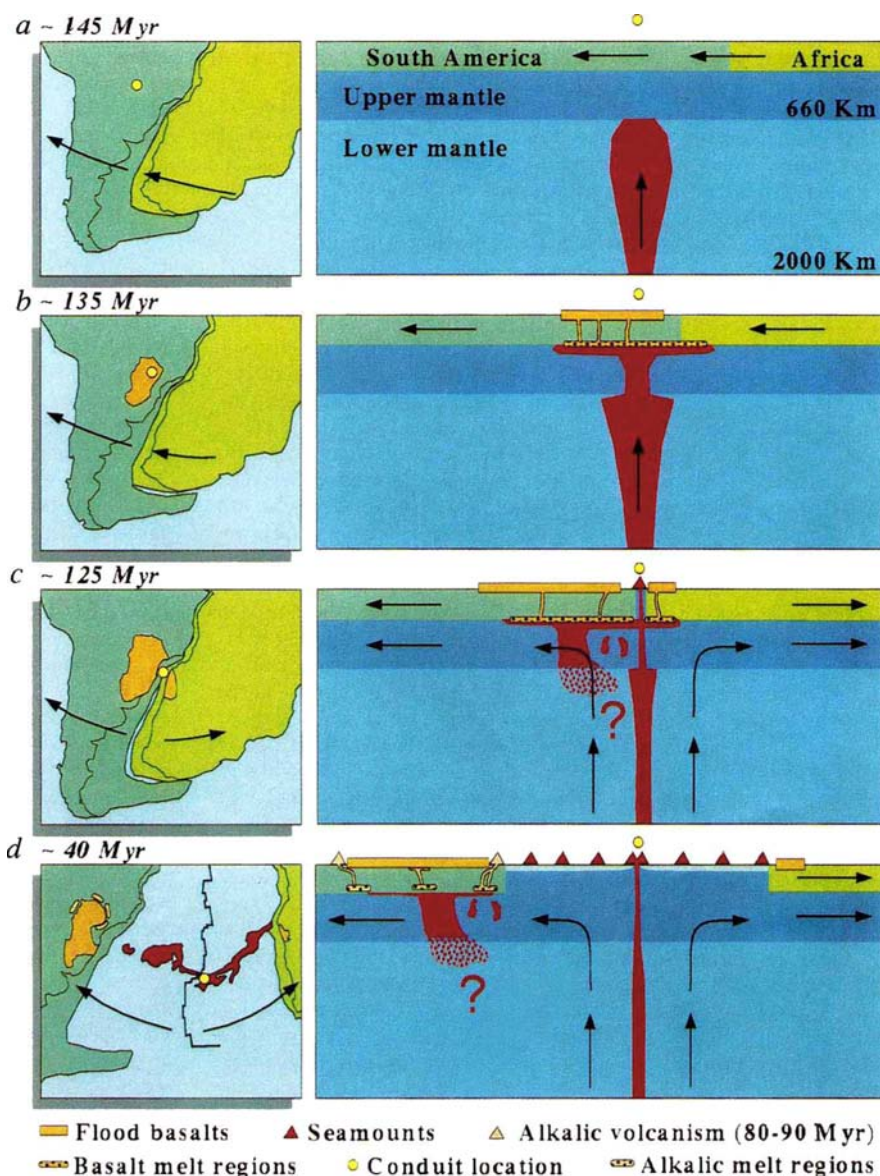
The P- and S-wave relative arrival times are independently inverted for compressional- and shear-wave slowness ($1/\text{velocity}$) perturbations as parametrized by the splines under tension. We invert simultaneously for the slowness perturbation field, earthquake relocations and station corrections. Although our data set is not highly dependent on earthquake location, an earthquake mis-location or heterogeneous structure far from our study area will induce a trend in time across the network. Although these trends tend to be inconsistent between events and therefore are not much mapped into structure, if unaccounted for they do bias our estimates of variance reduction used to construct pseudo-variance/resolution trade-off plots.

The station corrections are necessary to account for shallow crustal heterogeneity as well as differences in station elevations. The distributions of events used in this study (numbering 240 for P waves and 153 for S waves) provide effective 360° azimuthal coverage with respect to the Brazil network. The relative locations for events associated with ray paths that had turning points in the lower mantle are shown in Fig. 2. In addition to these, 58 of the events with P-wave arrivals and 40 of the events with S-wave arrivals resulted in 'core-phase' ray paths—where rays traversed Earth's core and impinged nearly vertically on the study area¹⁷. There were four to nine observations made for each of these events, leading to a total of 1,568 P-wave and 1,030 S-wave relative times to be inverted. Note that over this time period the coverage of events is fairly well distributed, in particular with respect to azimuth from the network. This is important, as the degree to which we have rays crossing through a region determines how well we can resolve structure there.

At this stage, our linear system is under-determined (all parameters are not uniquely determined by the data), so we must add some outside *a priori* information to obtain a unique solution or set of solutions. We follow the 'Occam's razor' approach to inversion, searching for the simplest model necessary—that is, the model that contains the least amount of structure required to satisfy the observations to within their estimated standard errors^{19–23}. We define 'structure' by model derivatives and seek to equally minimize spatial gradients and roughness implemented numerically through first- and second-difference operators.

We invert these large linear systems with a conjugate gradients procedure iterated to convergence²². We also iterate upon these inversions, systematically down-weighting equations associated

FIG. 6 Schematic diagram showing the principal stages of plume volcanism and upper-mantle flow beneath southeastern Brazil. Left-hand column, map views showing the breakup of Gondwanaland with the development of flood basalts and seamounts (after Turner *et al.*¹⁰). Right-hand column, a series of 'snapshots' of a cross-section beneath southeastern Brazil and western Africa, showing plume evolution. The plume head evolution is based on numerical modelling³⁰ of emplacement of a plume head from the lower mantle. *a*, Plume rises from the high-viscosity lower mantle. *b*, Plume ascends along a relatively narrow conduit through the low-viscosity upper mantle and mushrooms into highly flattened body 700 or 800 km in radius beneath the Brazil/Africa continental lithosphere. Melting of plume head material and/or hydrous lithospheric mantle produced the Paraná flood basalts over the time interval 137–127 Myr. *c*, Plume–continent relationship at 125 Myr, immediately before the onset of oceanic spreading between Africa and Brazil. Plume tail material is now emerging on the axis of the northward-propagating Mid-Atlantic Ridge. *d*, Relationship at 40 Myr. The Tristan da Cunha hotspot has produced substantial segments of the Walvis ridge and the Rio Grande rise. Late-stage volcanism is shown schematically around the margin of the Paraná basin, where small volume alkalic magmas pushed through the weaker lithosphere of the Brasília mobile belt.



with outlying residuals from the previous iteration^{20,24}. The effect of this is to produce a robust solution with L2 (least-squares) residual minimization within 1.5 residual standard deviations and, in the limit of many iterations, L1 (median) minimization for those equations associated with larger residuals. We perform 10 iterations of the robust down-weighting and 2,000 conjugate gradient iterations for each of these inversions. In the end, our models explain 86% of the root-mean-square data residual for P-wave data (from 0.29 to 0.04 s) and 91% for S-wave data (from 1.12 to 0.10 s), in line with our *a priori* estimates of data errors.

To illustrate the type of ray coverage present in our target region, and therefore the spatial resolution we can hope to obtain, we have plotted in Fig. 3a ray paths from 8 of the 240 events used in the P-wave inversion. Each colour represents a different event and the ray paths are shown leaving the region at 660 km depth—the underlying grey sphere. It is apparent from this figure why shallow structure (above ~80 km depth) will be difficult to resolve from station corrections because all of the rays sample the region beneath each station in a similar manner and there are no crossing paths available to sort out the ambiguity. Conversely, we see that at depths of 100–500 km the rays criss-cross one another at high angles, making it possible to

isolate anomalies in these regions. Outside the target region, the rays are once again travelling subparallel to one another making it difficult to map accurately the anomalies in these areas. It should be evident that with 30 times the number of events shown in Fig. 3a, the resolution of structure in the central region will be quite good. In Fig. 3b are plotted all rays which sample a small region at 250 km depth where we have observed the plume-like low-velocity anomaly. This figure demonstrates how the high density of converging rays makes it possible to isolate structure in this region.

A fossil plume beneath Brazil

The results of the nonlinear, robust, seismic travel-time inversions are shown in Figs 4 and 5. Figure 4 includes both constant-depth sections through the models and vertical cross-sections (shown as per cent velocity perturbation from a standard radial Earth model²⁵). The scales fade to black as our ability to resolve structure decreases, with pure black indicating that no rays traversed those regions. The most prominent and robust features of the travel-time inversions are: (1) the shallow high-velocity region beneath the São Francisco craton; (2) the deep high velocities to the southwest; and (3) the deep low-velocity 'plume-like' structure near the centre of the network. These features are

shown in perspective view in Fig. 5, where the volumes that enclose material with the highest seismic velocities are plotted in blue, and lowest seismic velocities are shown in red. The upper two plots in Fig. 5 show the view from directly above the study area, while the lower two plots are viewed from due south of the region. We have performed resolution tests (available on request: see Supplementary Information) which indicate that a deep low-velocity plume-like structure is well resolved by these data sets, whereas shallow structures (such as the high-velocity root beneath the São Francisco craton) are slightly underestimated due to the ability of the inversion to absorb some of the delay-time signal into the station correction parameters (the station corrections are damped so as to absorb only the approximate level of signal expected from very shallow structure). Although they have been obtained by entirely independent inversions, the P- and S-wave velocity models exhibit remarkably similar features, a strong indication that the structures are authentic. The absolute magnitudes of the structures are not extremely well determined—our resolution tests indicate that we should retrieve ~80% of the plume-like structure—but consistently it appears that the shear-wave anomaly is 1.5–2 times as strong as that of the compressional-wave anomaly.

The low-velocity anomaly, as found by our measurements, is roughly cylindrical in form, 300 km across and extending vertically from about 200 to at least 500–600 km depth (our resolution degrades severely at depths greater than 600 km). Three possible explanations for such an anomaly are that: (1) it represents a new plume rising through the upper mantle; (2) it is part of a small-scale convection cell; or (3) it is a relict related to the Paraná plume event. The first explanation is unlikely for several reasons. There is no indication of current tectonic activity above the anomaly or volcanic activity to indicate that a deep heat source is actively rising in the area. Also, the coincidence that a new plume happens to be ascending through the region where the Paraná plume rose 130 Myr ago seems highly implausible. The second explanation, that the anomaly is part of a small-scale upper-mantle convection cell, explains neither the morphological characteristics of the feature nor, once again, the lack of a surface expression which would be expected from a dynamically active convection cell. We therefore consider the third interpretation as the most plausible explanation for both our current observations of velocity structure and the evolution of the Paraná volcanism from 130–80 Myr.

We have interpreted the low-velocity feature shown in Figs 4 and 5 as the thermal, and possibly chemical, remnant of the original plume conduit that supplied the Paraná plume head. The shape and extent of the seismic anomaly is consistent with it being part of a plume head, or plume-head conduit, as observed in recent laboratory and numerical modelling experiments^{26–30}. These experiments show that a newly formed thermal plume develops a large 'head' and a much smaller 'tail'. The head is necessarily large by virtue of the buoyancy needed to displace overlying colder material, and the tail is simply maintained by buoyancy flux. The Paraná flood volcanism is generally attributed to a plume head, whereas the subsequent, less voluminous, volcanism associated with the Tristan da Cunha hotspot represents the plume tail. The region to the southeast of the cylindrical low-velocity anomaly where the plume-tail conduit is presumed to have cut through the upper mantle—in the opposite direction of absolute plate motion 130 Myr ago—is seen as a prominent low-velocity feature in the S-wave model, and somewhat more broken low-velocity feature in the P-wave model.

The features that could characterize a lower-mantle plume rising beneath a lithospheric plate have been described by Farneri and Richards³⁰. Figure 6 shows, in cartoon form, a plausible series of 'snapshots' of plume evolution beneath southeastern Brazil. The cross-sectional plume modelling is based on that used in ref. 30, and the volcanic and tectonic chronology is based on the summary by Turner *et al.*¹⁰. We assume here a viscosity contrast of about 1–2 orders of magnitude between the upper

and lower mantle, and between the upper mantle and the continental lithosphere. Thus, Fig. 6a shows the plume rising from high-viscosity lower mantle. In Fig. 6b, the plume ascends along a relatively narrow conduit through the lower-viscosity upper mantle and mushrooms beneath the Brazil/Africa continental lithosphere to a highly flattened body perhaps 700 or 800 km in radius. Melting of plume material and/or hydrous lithospheric mantle produced the Paraná flood basalts over the time interval 137–127 Myr. By 125 Myr (Fig. 6c), immediately before the onset of oceanic spreading between Africa and Brazil, material from the plume tail had begun erupting at Tristan da Cunha where it formed part of the axis of the northward-propagating Mid-Atlantic Ridge.

By 40 Myr (Fig. 6d), the Tristan da Cunha hotspot had produced substantial segments of the Walvis ridge and the Rio Grande rise. A major phase of late-stage alkalic volcanism occurred around the margin of the Paraná basin about 80–90 Myr ago, mostly within the Brasília mobile belt where lithosphere may have been thinner and/or weaker. During the entire time interval shown in Fig. 6, the position of the Paraná plume-head conduit is assumed to remain in close proximity to its original position in the upper mantle beneath the Paraná; that is, the lithosphere and the sub-lithospheric mantle are in coupled flow. A fixed position for the plume conduit with respect to the overlying continent is required in the above evolutionary sequence to explain the late-stage alkalic volcanism that borders the Paraná basin. This late Cretaceous volcanism seems to require a deep conductive heat source sufficient to melt the lower lithosphere at least 50–60 Myr after the main phase of basaltic volcanism had ended. There is little seismic evidence for the remains of the original plume head that presumably spread beneath much of the Paraná basin. The thermal signature may have been lost through extraction of heat (magma) during the Paraná flood volcanism. The shallower portions of our velocity models—in particular the S-wave model—are in general agreement with recent heat-flow measurements made in the region^{31,32} which indicate increasing heat flow from the central axis of the Paraná basin to the eastern margin and mobile belts.

If the observed seismic anomaly in the upper mantle is thermal in origin, we can make a rough estimate of the temperature excess based on laboratory measurements of changes in seismic velocity with respect to temperature³³. In the results presented above, the maximum contrast between 'normal' mantle and the centre of the plume feature is about 1.5% in P- and 2% in S-wave velocities. For a pyrolite mantle composition (57% olivine, 14% garnet, 17% orthopyroxene and 12% clinopyroxene), the observed velocity variations can be produced by a thermal anomaly of about 200 °C. This is at the low end of other estimates of excess temperature in the upper mantle of about 200–300 °C for the Hawaiian hotspot³⁴, based on petrogenetic constraints. The thermal signature of a mantle plume some 200–300 km in width will, where heat flow is purely conductive, persist for substantial periods of geological time. For a thermal diffusivity (the ratio of the thermal conductivity to the product of density and specific heat at constant volume) of $1 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$ in the upper mantle³⁵, a time period of 125 Myr will result in about a 40–50% decrease in centre temperature (suggesting an initial plume core temperature of about 350–400 °C). Over 125 Myr an initial plume 200 km wide will broaden only to about 260 km, not enough to alter significantly the shape or dimensions of the plume anomaly observed in the mantle beneath the Paraná basin. By the same token, the high-velocity mantle root, 200–300 km thick, that underlies the Archaean São Francisco craton has preserved its distinctive velocity (and presumably its thermal signature) over far longer periods of geological time. This indicates that, at least to these depths, the lithosphere and upper mantle have been moving coherently³. To some extent, the plume conduit is 'entrained' in this root structure, which may also help to explain the observation that the whole of the upper mantle in this region is moving as a unit.

The low-velocity anomaly that we observe need not be a purely thermal feature. If, for instance, the rising plume head contained compositionally distinct material from the lower mantle, or had become compositionally stratified—with increasing iron enrichment in the deeper parts of the plume, for example, then we might expect a compositionally induced velocity anomaly as well. Thus we may speculate that progressive iron enrichment of the material rising in the plume head would lead to higher densities that would act to compensate the thermally induced lower densities yet would enhance the low-velocity anomaly. The plausibility of this argument is based in part on Jordan's³⁶ calculations on the effect of Fe/Mg ratio on velocity in mantle rocks. For mantle rocks, an excess temperature of 200 °C results in a density decrease of about 0.6% (ref. 36). To compensate for this density decrease we require an approximately 10% increase in the FeO/(FeO + MgO) ratio. This percentage change is only about 20% of the range observed for continental mantle lherzolites, and about 50% of the contrast between average continental lherzolite and pyrolite. The calculated increase in Fe/Mg ratio would result in a velocity decrease of about 0.15% in P- and about 0.3% in S-wave velocities. If, in fact, a compositional gradient does exist it could lead to a natural end for advective flow, as the thermally induced buoyancy would drive flow in the plume only until such time that sufficient heat had been removed through advection to make the plume material neutrally buoyant. At this point only the much slower process of conduction would be available to dissipate the thermal anomaly—assuming no replenishment of fresh plume material from the deep mantle. Our knowledge of plume composition and compositional gradients is, of course, sufficient only to admit general plausibility arguments, but the balancing of thermal and

compositional density effects would provide a straightforward explanation of our observations. Thus the fact that volcanism in the region ceased at least 50 Myr ago would not be inconsistent with the continued presence of a deep plume residual, as advective heat transfer to the continental lithosphere would cease once the plume became neutrally buoyant.

Implications for plate driving forces

The results presented here have important implications for the mechanisms that drive continental plates^{37,38}. If our interpretation of the velocity structure as a fossil plume is correct, then the South American plate has been in coupled motion with its underlying sub-lithospheric mantle for at least 130 Myr. A recent force modelling study has shown that such coupling is consistent with a dynamical analysis of the South American plate⁷, but our results indicate that this coupling includes virtually the whole of the upper mantle—to depths of at least 500–600 km. Although this seems surprising, perhaps it should not be, for as South America moved to the northwest and then westwards, Nazca plate subduction also migrated progressively westward, in effect closing the South Pacific Ocean as the South Atlantic Ocean opens. If South America and its underlying upper mantle were not in coupled flow, additional material would be required to fill in that upper mantle, either from the lower mantle or a more remote part of the upper mantle. In the context of the opening of the South Atlantic, it would seem probable that new material is being fed from the region of the mid-ocean ridge and that the new material migrates westwards with the plate. Thus the movement of continental plates—whose keels may extend hundreds of kilometres into the mantle—may be controlled by convective overturn in the Earth's mantle to a far greater extent than previously recognized. □

Received 29 March; accepted 28 September 1995.

- Chapple, W. M. & Tullis, T. E. *J. geophys. Res.* **82**, 1967–1984 (1977).
- Sacks, I. S. & Okada, H. *Phys. Earth planet. Inter.* **9**, 211–219 (1974).
- Jordan, T. H. *Nature* **274**, 544–548 (1978).
- James, D. E., Assumpção, M., Snoke, J. A., Ribotta, L. C. & Kuehnel, R. *An. Acad. bras. Ci.* **65**, 227–250 (1993).
- Brito Neves, B. & Cordani, U. *Precamb. Res.* **53**, 23–40 (1991).
- Gripp, A. E. & Gordon, R. G. *Geophys. Res. Lett.* **17**, 1109–1112 (1990).
- Meijer, P. Th. & Wortel, M. J. R. *J. geophys. Res.* **97**, 11915–11931 (1992).
- Richards, M. A., Duncan, R. A. & Courtillot, V. E. *Science* **246**, 103–107 (1989).
- Gallagher, K. & Hawkesworth, C. J. *Nature* **358**, 57–59 (1992).
- Turner, S., Regelous, M., Kelley, S., Hawkesworth, C. J. & Mantovani, M. S. M. *Earth planet. Sci. Lett.* **121**, 333–348 (1994).
- O'Connor, J. M. & Duncan, R. A. *J. geophys. Res.* **95**, 17475–17502 (1990).
- Bizzi, L. A., Smith, C. B., de Wit, M. J., MacDonald, I. & Armstrong, R. A. *Int. Symp. Phys. Chem. Upper Mantle. Invited Lectures* (eds Teixeira, W. & Ribeiro, F. B.) 227–255 (Univ. São Paulo, 1994).
- Leonardos, O. H., Gibson, S. A. & Thompson, R. N. *Proc. Acad. Sci. India (W.S. Fife mem. Vol. (in the press))*.
- Gibson, S. A., Thompson, R. N., Leonardos, O. H., Dickinson, A. P. & Mitchell, J. G. *J. Petrology* (in the press).
- Toyoda, K., Horiuchi, H. & Tokonami, M. *Earth planet. Sci. Lett.* **126**, 315–331 (1994).
- VanDecar, J. C. & Crosson, R. S. *Bull. seism. Soc. Am.* **80**, 150–169 (1990).
- Aki, K. & Richards, P. G. *Quantitative Seismology—Theory and Methods* Vols 1 & 2 (Freeman, San Francisco, 1980).
- Neele, F., VanDecar, J. C. & Snieder, R. *Geophys. J. Int.* **115**, 482–496 (1993).
- Neele, F., VanDecar, J. C. & Snieder, R. *J. geophys. Res.* **98**, 12033–12054 (1993).
- Bostock, M. G. & VanDecar, J. C. *Can. J. Earth Sci.* **32**, 1–12 (1995).
- Constable, S. C., Parker, R. L. & Constable, C. G. *Geophysics* **52**, 289–300 (1987).
- VanDecar, J. C. & Snieder, R. *Geophysics* **59**, 818–829 (1994).
- Parker, R. L. *Geophysical Inverse Theory* (Princeton Univ. Press, 1994).
- Huber, P. J. *Robust Statistics* (Wiley, New York, 1981).
- Kennett, B. L. N. & Engdahl, E. R. *Geophys. J. Int.* **105**, 429–465 (1991).
- Olsen, P. & Singer, H. J. *Fluid Mech.* **158**, 511–531 (1985).
- Griffiths, R. W. & Campbell, I. H. *J. geophys. Res.* **96**, 18295–18310 (1991).
- Whitehead, J. A. & Luther, D. S. *J. geophys. Res.* **80**, 705–717 (1975).
- Manga, M., Stone, H. A. & O'Connell, R. J. *J. geophys. Res.* **98**, 19979–19990 (1993).
- Farnetani, C. G. & Richards, M. A. *J. geophys. Res.* **99**, 13813–13833 (1994).
- Hurter, S. J. thesis, Univ. Michigan (1992).
- Hurter, S. J. & Pollack, H. N. *Proc. Gen. Assembly Int. Union Geodesy Geophys. (IUGG)* (abstr.) A417 (Am. Geophys. Union, Washington DC, 1995).
- Christensen, N. I. *Handbook of Physical Properties of Rocks* Vol 2 (ed. Carmichael, R. S.) 1–228 (CRC, Boca Raton, 1982).
- Sleep, N. H. *J. geophys. Res.* **95**, 6715–6736 (1990).
- Turcotte, D. L. & Schubert, G. *Geodynamics* (Wiley, New York, 1982).
- Jordan, T. H. *The Mantle Sample* 1–14 (Am. Geophys. Union, Washington DC, 1979).
- Ziegler, P. A. *J. geol. Soc. Lond.* **150**, 927–940 (1993).
- Wilson, M. J. *J. geol. Soc. Lond.* **150**, 923–926 (1993).

SUPPLEMENTARY INFORMATION. Requests should be addressed to Mary Sheehan at the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. The Brazilian Lithosphere Seismic Project is a multinational collaborative effort between the Carnegie Institution of Washington and the Institute of Astronomy and Geophysics of the University of São Paulo. Other collaborators include Virginia Polytechnic Institute (A. Snoke), University of Illinois (T. Clarke), IPT-Brazil (L. Ribotta), and University of Montpellier II (A. Vauchez). We thank R. Kuehnel for station maintenance and data processing; R. Wortel for useful comments; and P. van Keken, H. Pollack, F. Press, S. Hurter, L. Solheim, R. Carlson, S. Solomon, and S. Sacks for stimulating discussions that focused our attention on various aspects of this problem. We have particularly benefitted from discussions of South American plate dynamics with R. Russo and P. Silver. This work was supported by the US NSF, FAPESP (Brazil) and PADCT (Brazil).

Structure of the fibre-forming protein pilin at 2.6 Å resolution

Hans E. Parge*, Katrina T. Forest, Michael J. Hickey, Deborah A. Christensen, Elizabeth D. Getzoff & John A. Tainer

Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA

The crystallographic structure of *Neisseria gonorrhoeae* pilin, which assembles into the multi-functional pilus adhesion and virulence factor, reveals an α - β roll fold with a striking 85 Å α -helical spine and an O-linked disaccharide. Key residues stabilize interactions that allow sequence hypervariability, responsible for pilin's celebrated antigenic variation, within disulphide region β -strands and connections. Pilin surface shape, hydrophobicity and sequence variation constrain pilus assembly to the packing of flat subunit faces against α_1 helices. Helical fibre assembly is postulated to form a core of coiled α_1 helices banded by β -sheet, leaving carbohydrate and hypervariable sequence regions exposed to solvent.

A MAJOR challenge in biology is to explain how proteins function as domains for the formation of larger assemblies critical to numerous cell processes. Fibre-forming proteins that form such assemblies as motor filaments, microtubules, flagella and pili are particularly difficult to study at atomic resolution, yet crystallographic structures of actin^{1,2} and myosin S1 fragment³ have had enormous impact on understanding motor filaments. Structural studies on tubulin⁴, flagella^{5,6} and uropathogenic P-pili⁷ have advanced our understanding of these assemblies but have not yet reached atomic resolution. Type-IV pili⁸, multifunctional virulence factors in many bacterial pathogens, are another class of fibres for which the progress in biochemical, genetic, immunological and biological studies would be enhanced by complementary structural information.

Type-IV pili are involved in adhesion to host cell surfaces^{9,10}, modulation of target cell specificity¹¹, twitching motility¹², bacteriophage adsorption and pilus retraction¹³. These immunogenic structures are required for colonization of the eukaryotic host¹⁴ by pathogenic bacteria including *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Pseudomonas aeruginosa*, *Dichelobacter nodosus*, *Moraxella bovis*, *Vibrio cholerae*¹⁵, and enterotoxigenic *Escherichia coli*¹⁶. Immunization with pilin prevents strain-specific infection by *N. gonorrhoeae* or *D. nodosus*^{17,18}, which cause gonorrhea and ovine foot rot, respectively. *N. gonorrhoeae* pilin, however, undergoes extensive sequence and antigenic variation^{19,21} as do the PfEMP1 antigens of malaria and the surface proteins of trypanosomes²². Characterization of how a protein structure can accommodate such extreme antigenic variation has implications for understanding protein folding and for the design of strain-independent vaccines.

These type-IV pili are 1,000–4,000-nm-long, 60 Å-diameter fibres formed by the ordered association of thousands of identical pilin subunits plus a few copies of pilus-associated proteins^{23,24}. Type-IV pili are assembled by a common cellular machinery¹⁵, share an unusual *N*-methyl phenylalanine, show high conservation of the N-terminal 32 amino acids²⁰, and have a proposed C-terminal immunogenic disulphide region. Several bacterial pilins are properly expressed and assembled in *Pseudomonas*^{25,26}, so assembly proteins for type-IV pili are conserved across species^{15,27}. In *Neisseria*, a tip-localized pilus-associated adhesin may mediate cell binding^{23,24}.

To understand the fibre-forming pilin protein and its assembly into pili, we solved the atomic resolution structure of a pilin

of relative molecular mass 17.4K from *N. gonorrhoeae* strain MS11. This structure provides the three-dimensional context for understanding antigenic variation, for mapping genetic, biochemical and immunological results, and for modelling pilus assembly.

Structure determination and refinement

To favour three-dimensional crystal growth over fibre formation, pilin crystals were grown from polyethylene glycol 400 (PEG400) in the presence of *n*-octyl- β -D-glucopyranoside (β OG) and 1,2,3-heptanetriol (Table 1 legend). The small, flexible, needle-shaped (0.5 mm $\times$ 0.05 mm $\times$ 0.05 mm) crystals have ~60% solvent content, belong to the orthorhombic space group C222, have cell dimensions $a = 127.6$, $b = 121.1$, $c = 26.9$ Å, and contain one molecule per asymmetric unit²⁸. The pilin structure was difficult to determine because of the small crystal size and the single major site for most of the heavy-atom derivatives used for multiple isomorphous replacement (MIR) plus anomalous scattering (MIRAS) phase determination. To maximize phasing information from anomalous scattering, diffraction data (Table 1) were collected from the $K_2Pt(NO_3)_4$ derivative with 1.07 Å X-radiation at the Stanford Synchrotron Radiation Laboratory (SSRL). Final electron density maps at 2.6 Å resolution clearly show positions for carbonyl oxygen atoms and side chains as well as the covalently bound disaccharide (Fig. 1). The atomic structure, consisting of 1,208 protein, 25 sugar, 129 water, 1 platinum, and 10 1,2,3-heptanetriol atoms, was refined to an *R*-factor of 19.4% for 2σ diffraction data in the 10–2.6 Å range, with root-mean square (r.m.s.) deviations on bond lengths and angles of 0.017 Å and 3.5°, respectively.

Pilin fold and structural chemistry

The elongated (85 $\times$ 34 $\times$ 26 Å), ladle-shaped pilin structure (Fig. 2a) is a novel version of an α - β roll fold composed of five elements (Fig. 2b): (1) the long hydrophobic α_1 -helix (residues 2–54); (2) the extended disaccharide-bound sugar loop (55–77); (3) two β -hairpins forming the four-stranded antiparallel β -sheet (78–93, 103–122) with all nearest neighbour (+1) β -strand connections; (4) the β_2 - β_3 loop connection (94–102); and (5) the disulphide region (121–158), which overlaps two residues with the β -sheet. The N-terminal segment (2–28) of the 85 Å long α -helical spine protrudes to form the ladle handle, whereas the C-terminal segment (29–54) cradled by the β -sheet creates the α - β roll forming the ladle head. Proteins sharing an α - β roll structure usually have related functions²⁹, so we have probably identified a new structural and functional class of α - β roll.

* Present address: Agouron Pharmaceuticals Inc., Research Laboratories, 3565 General Atomics Court, San Diego, California 92121-1121, USA.

TABLE 1 Diffraction data

Dataset*	Wavelength (Å)	Resolution (Å)	Completeness (%)	Unique reflections	$R_{\text{sym}}^{\dagger}$ (%)	$R_{\text{iso}}^{\ddagger}$ (%)	Soak (mM/days)	No. of sites	Phasing power§	$R_c^{\parallel}$ (%)
Native	1.07	2.6	95	6,494	8.1	—	—	—	—	—
Pt1	1.07	2.6	95	6,565	5.4	25.3	40/5	1	1.55/2.13	59/36
Pt2	1.08	3.3	95	3,342	12.0	26.6	20/13	2	1.52	59
Pt3	1.08	3.5	81	2,865	9.9	28.2	10/16	2	1.62	59
Pd1	1.07	3.0	94	4,236	13.3	30.8	0.5/1.5	1	1.28	62
Pb1	1.08	2.6	96	6,615	8.7	21.2	80/13	4	1.29	65
Pb2	1.07	3.0	81	3,641	8.3	16.3	10/1.5	2	1.14	68
Pb3	1.54	3.0	78	3,558	12.3	27.7	80/4	4	1.27	65
Pb4	1.54	3.1	90	3,749	16.4	33.5	80/7	4	1.23	65
Pb5	1.54	2.8	79	4,142	14.2	32.3	30/10	4	1.25	65
Pb6	1.54	2.8	83	4,353	15.6	28.0	40/18	4	1.25	61

Modified pilin purification and crystallization protocols²⁸, involving extensive filtration over an Amicon XM50 membrane and subsequent filtrate concentration to 20 mg ml⁻¹, permitted reproducible growth and macroseeding of needle crystals that diffract beyond 2.6 Å resolution. For derivatives, crystals were soaked in artificial mother liquor (60% PEG400, 50 mM CHES, pH 8.0, 1% β OG, 0.6% 1,2,3-heptanetriol) containing dissolved heavy-atom reagent. Heavy-atom sites were identified in difference-Patterson and difference-Fourier maps. The major platinum site at His 54 was occupied in all derivatives that left crystals intact, except for Me₃PbAc (which gave weak substitution at Glu 41, Glu 59 and Glu 82), dictating the need for anomalous phase information. The polypeptide chain was traced with XFIT⁴¹ at 3.0 Å resolution in MIRAS electron density maps that were improved (figure-of-merit increase from 0.75 to 0.86) by solvent flattening (with 50% solvent) using PHASES⁴². Missing segments were connected in maps calculated at 3.0 Å resolution with combined SIGMAA-weighted⁴³ phases from MIRAS and the partial model after simulated annealing refinement with XPLOR⁴⁴. Sequence assignment was guided by electron density for well defined landmark residues. Additional cycles of model building, refinement and phase combination at 2.6 Å resolution yielded a complete model that was then refined against the higher quality K₂Pt(NO₂)₄ dataset and refit to SIGMAA-weighted⁴³ simulated annealed omit maps. Crystal anisotropy, which resulted in weak diffraction along *a*, was corrected by applying an overall anisotropic *B* factor to *F_o* in the final stages of refinement. Residues Asn 154-Phe-Asp-Ala-Lys 158, which differ from the MS11 gene encoded sequence⁴⁵ Lys 154-Ala-Ser-Asp-Ala-Lys 159, were assigned from simulated annealed omit maps, have been reported in other pilins, and were confirmed by C-terminal sequencing (data not shown). Over 85% of the residues fall within the most favoured regions of the Ramachandran plot⁴⁶ and only turn residues Asn 85 and Asn 114 lie outside the allowed regions. Average *B* factors are 30 Å² for protein and 51 Å² for solvent.

* Datasets were collected at SSRL on a Mar Research image plate and in-house on a Siemens multiwire area detector, after screening of over 50 heavy-atom compounds by film data collection. Three heavy-atom derivatives were identified: Pt, K₂Pt(NO₂)₄; Pd, K₂Pd(Cl)₄; Pb, Me₃PbAc, where Me represents methyl and Ac acetyl groups.

† R_{sym} is the unweighted *R*-factor on *I* among symmetry mates.

‡ $R_{\text{iso}} = \Sigma(|F_{\text{ph}}| - |F_{\text{p}}|)/\Sigma|F_{\text{p}}|$, where $|F|$ is the structure-factor amplitude and the subscripts 'ph' and 'p' represent 'heavy-atom derivative' and 'native protein' respectively.

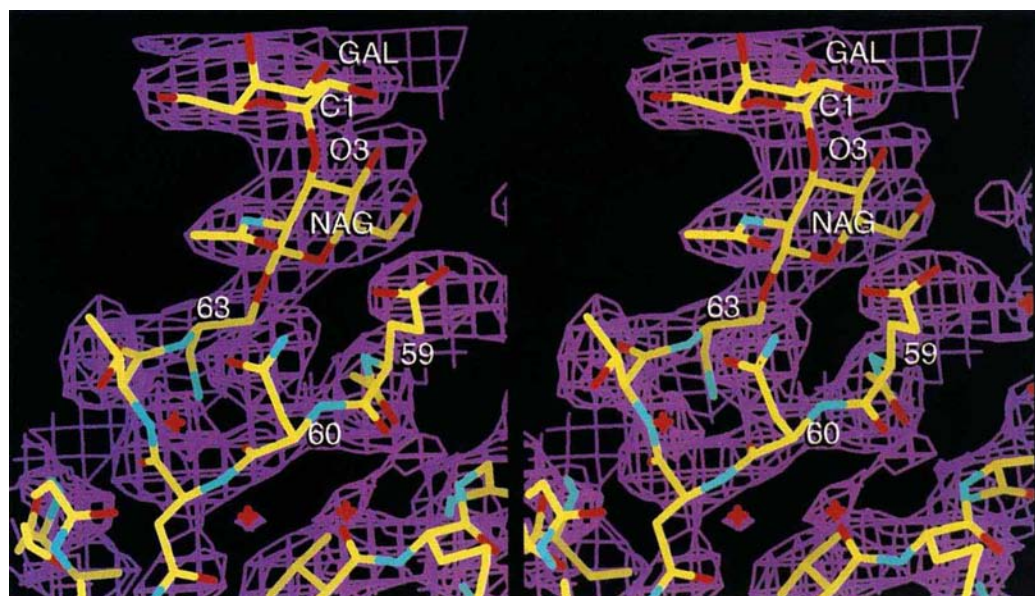
§ Phasing power is r.m.s. ($|F_{\text{h}}|/E$), where the subscript 'h' represents 'heavy-atom' and *E* is the residual lack of closure. The second entry for Pt1 is anomalous (ano) phasing power, r.m.s. ($|F_{\text{h}}^{\text{ano}}|/E_{\text{ano}}$).

|| $R_c = \Sigma||F_{\text{ph}}| \pm |F_{\text{p}}| - |F_{\text{h}}||/\Sigma||F_{\text{ph}}| \pm |F_{\text{p}}||$ for centric reflections. For the anomalous K₂Pt(NO₂)₄ dataset, R_c was calculated for the top 25% largest Bijvoet differences as $\Sigma||F_{\text{ph}}| - |F_{\text{phc}}|| + ||F_{\text{ph}}| - |F_{\text{phc}}||/\Sigma||F_{\text{ph}}| + |F_{\text{phc}}||$, where the subscript 'c' indicates 'calculated'.

The unusually long, 53-residue α_1 helix of pilin resembles the myosin S1 56-residue α -helix³. However, the pilin helix kinks at Pro 22 and Gly 42 to form an elongated S-curve, whereas the myosin helix gently curves. Three additional short helices occur within pilin loop connections: α_2 (60–64) and α_3 (69–72) in the sugar loop, and α_4 (99–102) in the β_2 – β_3 loop connection (Fig. 2).

Within the globular head domain, sequence-conserved residues (Fig. 2c) form two hydrophobic cores and link structural elements (Fig. 3). The larger hydrophobic core ties the C-terminal portion of α_1 to all four β -strands of the central sheet, as well as the β_2 – β_3 connection and the disulphide region. The smaller hydrophobic core between the β -sheet and disulphide region includes conserved residues from β_2 , β_3 , β_5 and β_6 . α_1

FIG. 1 Final pilin electron density map stereo view calculated at 2.6 Å resolution with $2F_o - F_c$ amplitudes and model phases, and contoured at 1 σ (purple). The refined model is coloured by atom type: C, yellow; N, blue; O, red; water molecules are shown as red crosses. The covalently bound disaccharide at Ser 63 is clearly defined as an O-linked GlcNAc- α 1,3-Gal. Asn 60 N-caps the short five residue α_2 helix (60–64) and stabilizes the disaccharide by hydrogen-bonding from Asn 60 N δ 2 to GlcNAc O7.



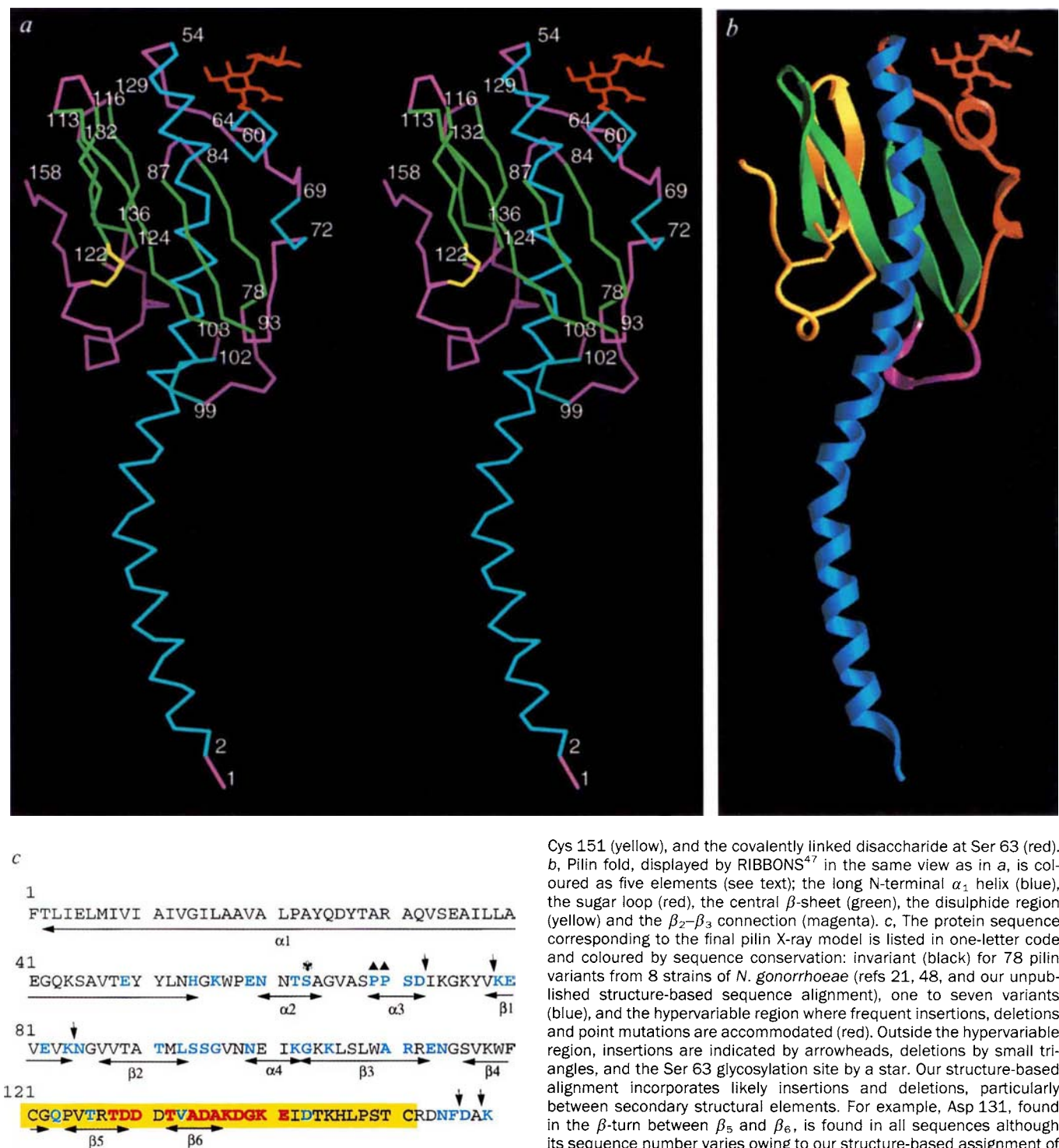


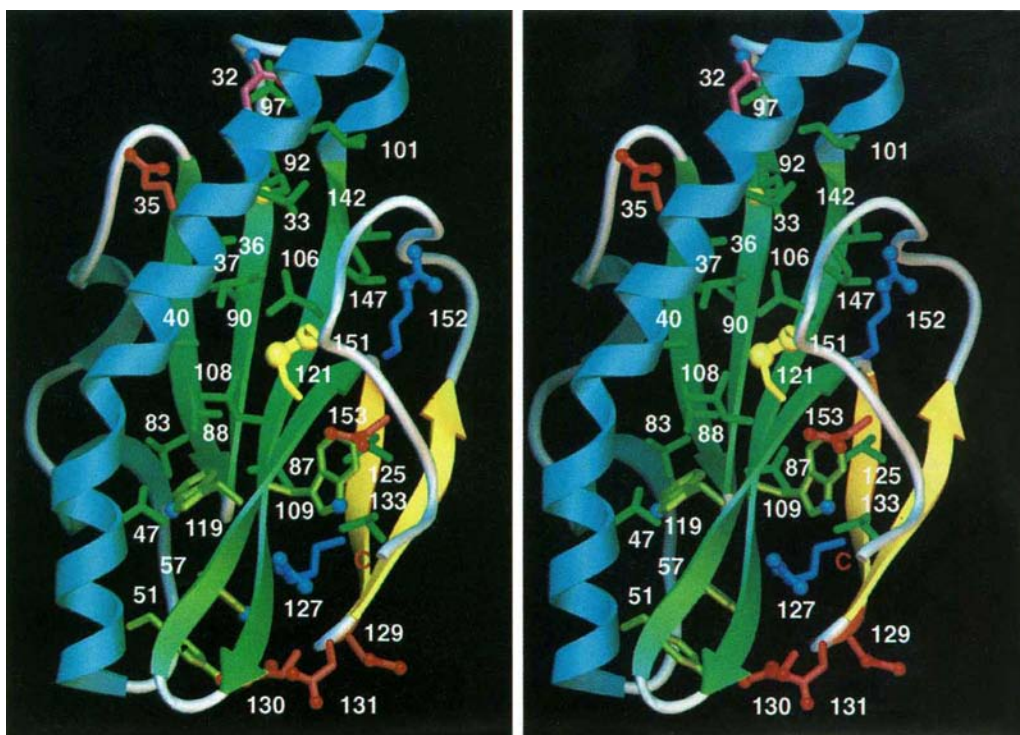
FIG. 2 Pilin fold, secondary structure and sequence variation. a, Pilin α -carbon backbone displayed in stereo with XFIT⁴¹. Residues at the beginnings and ends of secondary structural elements are labelled together with the N and C termini. Colours identify helices (blue), β -strands (green), coil (purple), the disulphide bond from Cys 121 to

Cys 151 (yellow), and the covalently linked disaccharide at Ser 63 (red). b, Pilin fold, displayed by RIBBONS⁴⁷ in the same view as in a, is coloured as five elements (see text); the long N-terminal α_1 helix (blue), the sugar loop (red), the central β -sheet (green), the disulphide region (yellow) and the β_2 - β_3 connection (magenta). c, The protein sequence corresponding to the final pilin X-ray model is listed in one-letter code and coloured by sequence conservation: invariant (black) for 78 pilin variants from 8 strains of *N. gonorrhoeae* (refs 21, 48, and our unpublished structure-based sequence alignment), one to seven variants (blue), and the hypervariable region where frequent insertions, deletions and point mutations are accommodated (red). Outside the hypervariable region, insertions are indicated by arrowheads, deletions by small triangles, and the Ser 63 glycosylation site by a star. Our structure-based alignment incorporates likely insertions and deletions, particularly between secondary structural elements. For example, Asp 131, found in the β -turn between β_5 and β_6 , is found in all sequences although its sequence number varies owing to our structure-based assignment of insertions and deletions near this exposed turn within the hypervariable region. Helices and sheets are underlined with arrows: α_1 2-54, α_2 60-64, α_3 69-72, α_4 99-102, β_1 78-84, β_2 87-93, β_3 103-113, β_4 116-122, β_5 124-129 and β_6 132-136. The structure-based sequence differs from the MS11 pilin sequence⁴⁵ near the C terminus (Table 1 legend).

is tied by side-chain-to-main-chain hydrogen bonds from Gln 32 to the β_2 - β_3 loop and from Glu 35 to the sugar loop. One conserved residue from the extended C-terminal tail, Arg 152, nucleates a connecting loop to stabilize disulphide bond formation, and two others connect to the β -sheet, one by hydrophobic interactions (Ala 157 to Phe 120) and the other by a side-chain-

to-main-chain hydrogen bond (Asp 153 to Cys 121) (Fig. 3). The most conserved link is the disulphide bond from C-terminal tail residue Cys 151 to β_4 Cys 121 (Fig. 2). These cysteines occur in the C-terminal third of all type-IV pilins¹⁹. Although residues flanking the disulphide cysteines are conserved among *N. gonorrhoeae*, intervening residues 128-141 vary considerably

FIG. 3 Stereo view of the globular pilin head domain with helices in cyan, central β -sheet in green, disulphide β -strands in yellow, and loops in white, as displayed using a local RIBBONS⁴⁷ module of AVS (Waltham, MA). Conserved hydrophobic side chains forming the core regions (green) and conserved polar side chains linking structural elements (red for negative, blue for positive and pink for neutral) are shown with polar side-chain atoms as spheres (red for oxygen, blue for nitrogen and yellow for sulphur). Referral to the secondary structure and sequence shown in Fig. 2c may aid interpretation. The smaller C-terminal core between the sequence-hypervariable disulphide region and the central β -sheet involves β_2 Val 87, β_3 Trp 109, β_5 Val 125 and β_6 Val 133. The more extensive core between the central sheet and α_1 involves α_1 Val 33, Ala 36, Ile 37, Ala 40, Val 47, Tyr 51; β_1 Val 78, Val 81 (both hidden behind α_1 and not labelled), Val 83; β_2 Val 88, Ala 90, Met 92; β_3 Leu 106, Leu 108; β_4 Trp 119, Cys 121; β_2 - β_3 connection Val 97, Ile 101; and disulphide region Ile 142, Leu 147, Cys 151. Conserved β_4 Phe 120 (not shown) connects to C-terminal tail Ala 157 (not shown) through hydrophobic contacts. Functionally conserved residues form side-chain-to-main-chain hydrogen bonds that link the structural elements. Gln 32 O ϵ 1 links α_1 to the β_2 - β_3 connection Asn 98 N. Glu 35 O ϵ 1 and O ϵ 2 form bifurcated hydrogen bonds: O ϵ 2 tying α_1 to sugar loop Lys 76 N and Tyr 77 N, and O ϵ 1 tying α_1 to sugar loop Tyr 77 N and β_1 Val 78 N. Arg 127 links the disulphide region to the central β -sheet: N ϵ to Asn 85 O, N η 2 to Ala 110 O, and N η 1 to Asp 129



(Fig. 2c) to allow escape from the human immune response, and have been termed the 'hypervariable region' within the 'disulphide loop'^{19,30}.

The pilin structure surprisingly reveals that this disulphide region is not an unstructured loop as implied by its sequence insertions and deletions, but a regular β -hairpin (β_5 - β_6) followed by a loop connection (137-150) to the extended C-terminal tail (151-158). Conserved residues Arg 127 and Ile 142 demarcate the hypervariable portion of the disulphide region (Fig. 3), and conserved Val 133 is centrally located within it (Fig. 2c). The β_5 Arg 127 guanidinium group interconnects the three β -hairpins with three hydrogen bonds, and β_5 with the sugar loop by aromatic stacking with Trp 57, which in turn stacks with α_1 Tyr 51 (Fig. 3). Ile 142 holds the C-terminal loop connection to the globular head by hydrophobic packing. Val 133 is part of the small hydrophobic core which tethers the central β -sheet to the disulphide β -hairpin. These three residues allow the two disulphide region β -strands and the first half of the C-terminal loop connection to vary in length and sequence by providing key conserved links that hold the disulphide region against the rest of the structure.

O-Linked disaccharide

At Ser 63, the electron density matches a covalently bound, O-linked, N-acetyl glucosamine- α 1,3-galactose (GlcNAc- α 1,3-Gal) (Fig. 1). Consistent with this carbohydrate structure, antibodies that bind the trisaccharide Gal- α 1,3-Gal- β 1,4-GlcNAc recognize *N. meningitidis* pilin³¹. N-glycosylation of Asn 60 within the eukaryotic N-glycosylation motif Asn 60-Asn-Thr-

O. Arg 152 anchors the C-terminal tail to the C-terminal loop connection: N ϵ and N η 2 to Leu 147 O, and N η 2 to Thr 144 O. These hydrogen bonds stabilize the disulphide loop connection for Cys 151 disulphide bond formation to Cys 121. Asp 153 links the C-terminal tail to the central β -sheet: O δ 1 to Cys 121 N. There is one main-chain-to-main-chain hydrogen bond from β_4 to the C-terminal tail: Cys 121 O to Asp 153 N. Sugar loop Trp 57 bridges the smaller disulphide and larger central cores by stacking with α_1 Tyr 51 and β_5 Arg 127. Asp 129, 130 and 131 'stutter' in our sequence alignment and form the β -hairpin in the disulphide region.

Ser 63 of *N. meningitidis* pilin was proposed, based upon presence of carbohydrate in peptide 45-73 and decreased epithelial cell adherence of a strain with pilin variant Asp 60-Asn-Ser-Ser 63³². In our pilin structure, the first four α_2 residues (Asn 60-Asn-Thr-Ser 63) also match the N-glycosylation motif, but the electron density clearly indicates an O-linked sugar, unprecedented in prokaryotic proteins. The Asn 60 side chain does N-cap this helix and hydrogen-bond to the disaccharide (Fig. 1). Thus, Asn 60 mutation might increase glycosylation site flexibility, whereas observed pilin variants at Ser 63 (Fig. 2c) are predicted not to be glycosylated or to be glycosylated elsewhere, providing antigenic variation and possible modulation of target-cell specificity and adhesivity.

Coiled-coil dimer

Pilin subunits assemble into twofold-symmetric dumb-bell-shaped dimers (123 Å × 31 Å × 22 Å) through the unusual anti-parallel coiled-coil packing of hydrophobic α_1 residues 3-25 (Fig. 4a, b). In the crystal lattice, pilin dimers pack into flat planes separated by the short unit-cell dimension (c = 26.9 Å). An extended 1,2,3-heptanetriol molecule aligned with this c -axis bridges symmetry-related hydrophobic α_1 helices near the N-terminal N-methylphenylalanine. The adjacent crystal lattice channel, surrounded by hydrophobic sides of the globular heads near α_1 Leu 38, probably contains the detergent β OG. In solution, β OG disassembles type-IV pilus fibres into pilin dimers^{28,33} which each bind about 30 β OG molecules³³. For *N. gonorrhoeae* pilin, dimer formation buries 1,051 Å² of hydrophobic surface area (Fig. 4b). The large buried surface area, absolute sequence

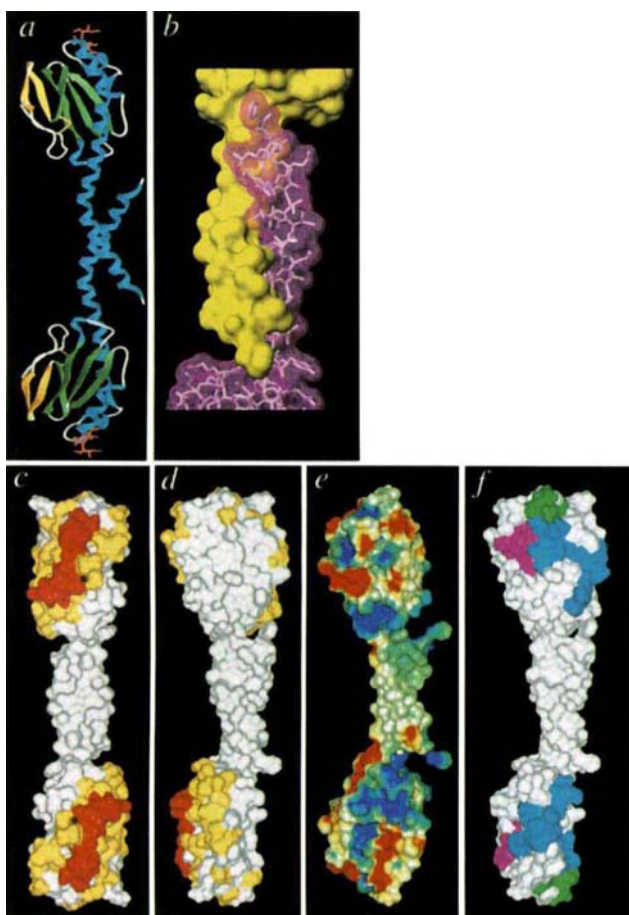


FIG. 4 Pilin dimer and molecular surface properties. **a**, Coiled-coil dimer assembly, viewed down the crystallographic *c* axis. The 15-turn α_1 helices (blue) cross at the two-fold crystallographic *a*-axis (horizontal), placing the two head domains (β_1 , β_2 , β_3 and β_4 in green, loops in white, disulphide region strands β_5 and β_6 in yellow and disaccharide in red) at opposite ends of the dimer. **b**, Intimate, highly complementary dimer interface, viewed at $\sim 90^\circ$ to the view in **a**, with N-terminal phenylalanines in front. Hydrophobic residues Ile 4, Met 7, Ile 10, Leu 21 and Tyr 24 each contribute $>40 \text{ \AA}^2$ of buried surface area. Even Gly 14 on the 2-fold axis contributes $37 \text{ \AA}^2$. **c**, Absolute sequence conservation (grey) and variation mapped to the molecular surface viewed at $\sim 180^\circ$ from view in **b**. Variable (yellow) and hypervariable (red) residues among 78 *N. gonorrhoeae* variants (Fig. 2c) cluster on the rounded face of the pilin head. **d**, Pilin dimer as in **c**, viewed 45° away. The hydrophobic flat face of the head domain (top) and the coiled-coil (centre) are invariant (grey). **e**, Electrostatic potential surface, view as in **a**. Lys residues 74, 76, 79, 102, 104, 105 and 140 form the dominant positive (blue) patch (top of lower head). Arg 30, Lys 137, Lys 145 and Arg 152 continue the positive band around the neck of the globular head (bottom of upper head). This band is broken by negatively charged (red) Asp 26, Glu 100, Asp 138, Glu 141 and Asp 143 (top left of lower head). Maximum colour saturation of blue and red corresponds to potentials of $\geq |2 \text{ kT/e}|$. **f**, Epitope mapping, view as in **d**. Peptides that elicit antibodies blocking cell binding (blue, residues 41–50 and 64–89)³⁷ are near clustered epitopes for crossreactive monoclonal antibodies that recognize all *N. gonorrhoeae* and a subclass of *N. meningitidis* pilins: SM1 (green, 49–53; ref. 36), SM2 (magenta, 109–128; ref. 36) and 7BE11 (subset of SM2, 117–124; our unpublished results).

METHODS. Surface area buried in the dimer interface was calculated with MS⁴⁹. Poisson–Boltzman electrostatic potential was calculated with DelPhi⁵⁰ as implemented in InsightII (Biosym Technologies) by using an ionic strength of 145 mM, probe radius of 1.4 Å, ion-exclusion radius of 2 Å, and dielectric constants of 2 and 80 for the protein and solvent, respectively. United-atom Discover CFF91 partial charges assuming a pH of 7 were used for the protein. For Ser 63 and its disaccharide, mulliken charges were calculated with the MOPAC module in Insight II. The $60 \times 47 \times 78 \text{ \AA}$ grid (with 1.5 Å grid spacings) centred on the dimer provided a 10 Å border on all sides. A full coulombic treatment was used at the grid boundaries.

conservation, and high complementarity of pilus dimer packing are of unknown relevance, but dimer building blocks for type-IV pilus fibres were proposed from biophysical studies³⁴.

Molecular surface properties

A striking sequence-conserved surface on the flat face of the α - β globular head (Fig. 4c, d) extends the entire length of the α_1 helical spine (Fig. 4d). This surface is composed of α_1 , the α_3 - β_1 connection, β_4 near Trp 119, and the disulphide region near Thr 150. Surface variability, including the river of hypervariable residues, localizes to the opposite rounded face of the globular head (Fig. 4c). Surface regions sequence-conserved among gonococcal pilins are implicated in species-specific functions such as

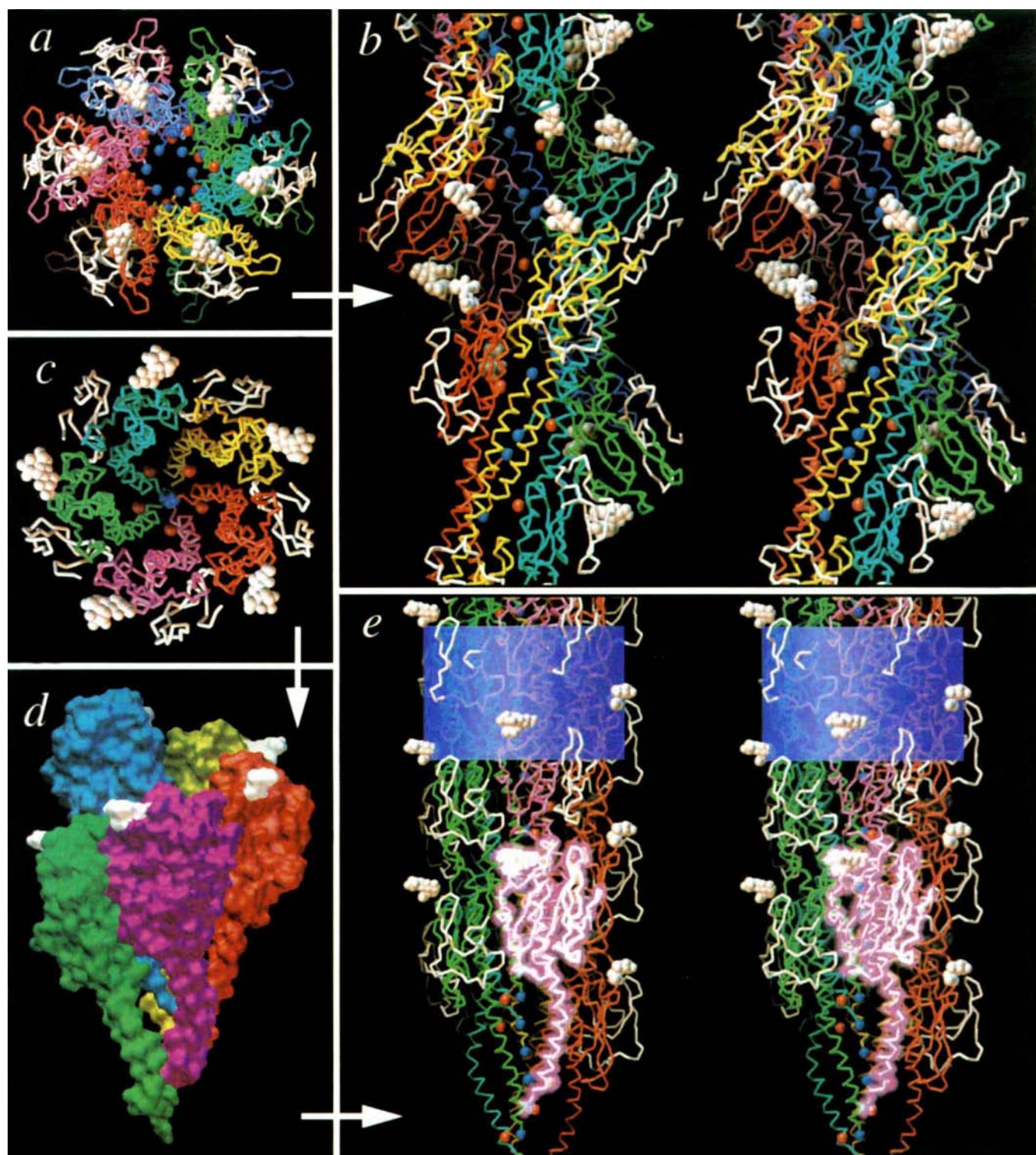
target cell recognition and accessory protein binding. Complete conservation across all type-IV pilin species, such as observed in the α_1 helix, argues for a fundamentally conserved function, namely fibre formation.

Positive electrostatic surface patches (Fig. 4e) may form non-specific DNA-binding sites as piliation is linked to transformation competence³⁵. Both positive and negative patches (Fig. 4e) may act in formation of pili bundles at physiological pH and ionic strength. Dissociation of bundles into single, soluble fibres at pH 9–10 is consistent with electrostatic fibre-fibre contacts. Conserved surfaces wrapping from the tip to the flat face of the globular head domain form contiguous epitopes for crossreactive monoclonal antibodies, raised against *N. meningitidis* pilin

FIG. 5 Dimer- and monomer-based fibre assemblies. **a**, A relatively high-scoring dimer-based sixfold helical model, generated as described below, viewed down the fibre axis. Six repeating dimers are shown as α tubes coloured red, yellow, cyan, green, blue and purple, with the disulphide region (121–158) white. The disaccharide (white), Phe 1 (blue) and Glu 5 (red) are shown as spheres. **b**, Stereo view of the model shown in **a**, with the fibre axis vertical. Hypervariable residues (white) are exposed on a ridge between a shallow groove, in which some conserved α_1 residues are exposed, and a deep groove, towards which the sugars face. The pitch (120 Å) and diameter (93 Å) are much larger than expected, the positively charged N termini of neighbouring dimers are close, and the fibre appears unexpectedly striated. **c**, A high-scoring monomer-based fivefold helical model generated and ranked as described below, viewed down the fibre axis looking into the conical opening at one end of the fibre, with five repeating monomers in green, purple, red, yellow and blue. Disulphide regions and sugars (white) are exposed at the periphery. **d**, A single helical turn of the monomer-based

fibre model shown in **c**, viewed with the fibre axis vertical, resembles a waffle ice-cream cone. Solid molecular surfaces (using the same colour scheme as in **c**) emphasize the continuous spiralling conical packing of pilin subunits within the fibre. **e**, Stereo view of the monomer-based fibre model in **c** and **d**, oriented as in **d**. A single molecule is highlighted to show how it packs into the fibre. The transparent blue ring is a molecular ruler showing that the model dimensions match the experimentally derived 41 Å pitch and 60 Å diameter. The continuous non-striated appearance matches electron microscopy results³⁹.

METHODS. To test whether the pilin structure can be assembled into pilin fibres with appropriate dimensions, 17,000 dimer-based and 99,000 monomer-based fibre models were generated by computation. For each model, an initial pilin module was oriented by three rotation angles, positioned a radial distance from the fibre axis, and systematically replicated by a translation defining the helical rise and a rotation defining the number of molecules per helical turn (twist). Rotation angles r_x (rotation about the subunit x-axis, which as the first rotation



applied, rotates the head of the subunit out of the fibre), r_y (rotation about the subunit axis originally normal to the fibre surface, which itself has rotated with r_x), r_z (rotation about the subunit long axis), rise, twist and radius were allowed to vary from $-10^\circ \rightarrow 5^\circ$, $-20^\circ \rightarrow 30^\circ$, $-90^\circ \rightarrow 90^\circ$, $8 \rightarrow 15$ Å, $6^\circ \rightarrow 124^\circ$ (plus -72°), and $-16 \rightarrow -32$ Å, respectively, for monomer searches, and from $0^\circ \rightarrow 0^\circ$, $-60^\circ \rightarrow 60^\circ$, $-10^\circ \rightarrow 10^\circ$, $8 \rightarrow 24$ Å, $-160^\circ \rightarrow 120^\circ$ and $-16 \rightarrow -24$ Å, respectively, for dimer searches. These search space limits for computational modelling were chosen based on initial graphical inspection of dimer and monomer models using much broader parameters. The following criteria were used to reject and rank models computationally before manual optimization using the graphics visualization software AVS: proximity of conserved negative side chains to conserved positive side chains or to the charged N terminus (models were rejected if there were

not at least two such interactions using Glu 5, Glu 41, Glu 49 C δ to Arg 30 C ζ , Lys 44 N ζ , Lys 76 N ζ , or the N terminus); burial of conserved hydrophobic side-chain atoms (model was rejected if the most distal side-chain carbon was farther than 6 Å from another conserved hydrophobic side chain for 8 or more residues); number of intersubunit $\text{Ca}-\text{Ca}$ distances less than 3.0 Å (models were rejected if there were more than 12) or less than 1.4 Å (models were rejected if there were any). Models were also rejected if the maximum $\text{Ca}-\text{Ca}$ diameter was greater than 84 Å. For all dimer models, these criteria were too stringent and were relaxed. Helix flexibility and packing of internal hydrophobic side chains, which might affect diffraction results and/or uptake of negative stain in electron microscopy, have not yet been addressed. Local AVS modules and computational assistance for fibre modelling were provided by T. J. Macke and M. E. Pique.

but also reactive against *N. gonorrhoeae* pilin³⁶, and peptide antigens for antipeptide antibodies that block cell binding³⁷ (Fig. 4f).

Implications for fibre assembly

To explore fibre assembly, we generated over 115,000 models by systematically building repetitive right- and left-handed helical assemblies of pilin monomers and dimers. We used steric collisions, charge and hydrophobic complementarity within the fibre contacts, minimal exposure of α_1 hydrophobic residues, maximal exposure of carbohydrate and hypervariable sequences, and helical parameters from fibre diffraction and electron microscopy as a computational basis for evaluation. Fibre diffraction on the ~15K *P. aeruginosa* PAK and PAO pili indicates that type-IV bacterial pili have five subunits per turn of helix, 40–41 Å pitch, 62 Å outer diameter, and subunit α -helices approximately parallel to the fibre axis^{34,38}. Electron micrographs of frozen-hydrated *N. gonorrhoeae* pili indicate a diameter of about 70 Å (data not shown). Sequence variation, elongated subunit shape, and fibre diffraction results severely limited possible packing arrangements, and defined boundaries for a search space within which the α_1 helix remains roughly parallel to the fibre axis, the fibre diameter is not too large, and conserved residues face inward. These constraints are best satisfied by packing the sequence-conserved flat face of the ladle head of one pilin subunit against the α_1 helical tail of another subunit in the fibre (Fig. 5).

Dimer and monomer building blocks produce strikingly different fibre models (Fig. 5). Dimer assembly models did not simultaneously satisfy all criteria for native pilus fibres: even the best dimer examples had diameters over 50% greater than expected, large tilts of the α_1 helix relative to the fibre axis, only partially exposed sugar moieties, and a clearly striated appearance (Fig. 5a, b). These computational results agree with experimental data from detergent-solubilized dimers, which retain native secondary structure but reassemble to form striated fibres with about twice the normal diameter³⁹. Dimers may occur naturally as membrane-associated assembly or disassembly forms of pilin that merit further study, or alternatively, may form only to maximally associate hydrophobic regions of α_1 helices during disassembly in detergent.

Monomer assembly models consistent with fibre diffraction and electron microscopy results produce polar fibres with appropriate pitch and diameter, and only the sugar and hyper-variable regions protruding. The β -sheet and sugar loop form a continuous 25-stranded tubular β -wrap around a core of packed α_1 helices in a plausible right-handed fibre model with five subunits per turn, a pitch of 41 Å, and an approximate diameter of 60 Å (Fig. 5c, d, e). This construction, a feasible assembly mode with no unresolvable collisions between subunits, could provide the high mechanical stability required for a fibre whose length (40,000 Å) is over 600 times its diameter (60 Å). The two α_1 helix kinks allow parallel coiled-coil helix packing, with a potential salt bridge from the positively charged N terminus of subunit^{N+1} to the conserved negatively charged side chain from Glu 5 of subunit^N, providing a mechanism for helix registration. In *P. aeruginosa*, Glu 5→Lys pilin incorporates into heterogeneous fibres with a corkscrew morphology, but does not form homogeneous fibres⁴⁰, supporting this registration role for Glu 5 in assembly. Left-handed helical fibre models expose conserved molecular surface and place the twisted β -sheet at an angle preventing continuous interactions.

Our ability to readily construct pilus fibre models consistent both with the highly asymmetric subunit and molecular surface properties of *N. gonorrhoeae* pilin, and with the biophysical characteristics of native *P. aeruginosa* pilus fibres, supports structural similarity among the family of type-IV pilins. In the resulting pilus models, carbohydrate and hypervariable regions protruding from a smooth cylinder (Fig. 5c) could be both bacterial cloaking devices against the host immune responses and modulators of pilus-based cell targeting. Three-dimensional electron microscopy reconstruction, such as recently accomplished for the *E. coli* Pap pili⁷, and structure-based site-directed mutagenesis promise to allow the experimentally guided assembly of the pilin structure into an atomic resolution model for the type-IV pilus. Such increased understanding of this highly versatile virulence factor should aid drug and vaccine design, with wide-ranging applications to the many bacterial pathogens that require type-IV pili for host cell targeting and infection. □

Received 18 July; accepted 27 September 1995.

- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. *Nature* **347**, 37–43 (1990).
- Schutt, C. E., Myslik, J. C., Rozycki, M. D., Goonasekera, N. C. W. & Lindberg, U. *Nature* **365**, 810–816 (1993).
- Rayment, I. et al. *Science* **261**, 50–58 (1993).
- Nogales, E., Wolf, S. G., Khan, I. A., Luodena, R. F. & Downing, K. H. *Nature* **375**, 424–427 (1995).
- Mimori, Y. et al. *J. molec. Biol.* **249**, 69–87 (1995).
- Morgan, D. G., Owen, C., Melanson, L. A. & DeRosier, D. J. *J. molec. Biol.* **249**, 88–110 (1995).
- Bullitt, E. & Makowski, L. *Nature* **373**, 164–167 (1995).
- Ottow, J. C. G. A. *Rev. Microbiol.* **29**, 79–108 (1975).
- Woods, D. E., Strauss, D. C., Waldermar, G. J. Jr, Berry, V. K. & Bass, J. A. *Infect. Immun.* **29**, 1146–1151 (1980).
- Buchanan, T. M. & Pierce, W. A. *Infect. Immun.* **13**, 1483–1489 (1976).
- Jonsson, A., Ilver, D., Falk, P. & Normark, S. *Molec. Microbiol.* **13**, 403–416 (1994).
- Bradley, D. E. *Can. J. Microbiol.* **26**, 146–154 (1980).
- Bradley, D. E. *Virology* **58**, 149–163 (1974).
- Dalrymple, B. & Mattick, J. S. *J. molec. Evol.* **25**, 261–269 (1987).
- Strom, M. S., Nunn, D. N. & Lory, S. *Meth. Enzym.* **235**, 527–540 (1994).
- Taniguchi, T., Fujino, Y., Yamamoto, K., Miwatani, T. & Honda, T. *Infect. Immun.* **63**, 724–728 (1995).
- Boslego, J. W. et al. *Vaccine* **9**, 154–162 (1991).
- Stewart, D. J., Clark, B. L., Peterson, J. E., Griffiths, D. A. & Smith, E. F. *Res. vet. Sci.* **32**, 140–147 (1982).
- Schoolnik, G. K. & Mietzner, T. A. in *New Generation Vaccines* (eds Woodrow, G. C. & Levine, M. M.) 565–597 (Dekker, New York, 1990).
- Mattick, J. S., Hobbs, M., Cox, P. T. & Dalrymple, B. P. in *Genetics and Molecular Biology of Anaerobic Bacteria* 517–545 (Springer, New York, 1993).
- Seifert, H., Wright, C., Jerse, A., Cohen, M. & Cannon, J. J. *clin. Invest.* **93**, 2744–2749 (1994).
- Borst, P., Bitter, W., McCulloch, R., Van Leeuwen, F. & Rudenko, G. *Cell* **82**, 1–4 (1995).
- Rudel, T., Scheuerpflug, I. & Meyer, T. F. *Nature* **373**, 357–359 (1995).
- Nassif, X. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 3769–3773 (1994).
- Mattick, J. S. et al. *J. Bact.* **169**, 33–41 (1987).
- Elleman, T. C. & Peterson, J. E. *Molec. Microbiol.* **1**, 377–380 (1987).
- Dupuy, B. & Pugsley, A. P. *J. Bact.* **176**, 1323–1331 (1994).
- Parge, H. E. et al. *J. biol. Chem.* **265**, 2278–2285 (1990).

- Orengo, C. & Thornton, J. M. *Structure* **1**, 105–120 (1993).
- Seifert, H. & So, M. *Microbiol. Rev.* **52**, 327–336 (1984).
- Hamadeh, R. M., Jarvis, G. A., Estabrook, M. & Griffiss, J. M. in *Ninth Int. Pathogenic Neisseria Conf.* 132–133 (SCC, Winchester, 1994).
- Virji, M. et al. *Molec. Microbiol.* **10**, 1013–1028 (1993).
- Watts, T. H., Kay, C. M. & Paranchych, W. *Can. J. Biochem.* **60**, 867–872 (1982).
- Watts, T. H., Kay, C. M. & Paranchych, W. *Biochemistry* **22**, 3640–3646 (1983).
- Biswas, G. D., Sox, T., Blackman, E. & Sparling, P. F. *J. Bact.* **129**, 983–992 (1977).
- Virji, M., Heckels, J. E., Potts, W. J., Hart, C. A. & Saunders, J. R. *J. gen. Microbiol.* **135**, 3239–3251 (1989).
- Rothbard, J. B., Fernandez, R., Wang, L., Teng, N. N. H. & Schoolnik, G. K. *Proc. natn. Acad. Sci. U.S.A.* **82**, 915–919 (1985).
- Folkhard, W., Marvin, D. A., Watts, T. H. & Paranchych, W. *J. molec. Biol.* **149**, 79–93 (1981).
- Watts, T. H., Scraba, D. G. & Paranchych, W. *J. Bact.* **151**, 1508–1513 (1982).
- Pasloske, B. L., Scraba, D. G. & Paranchych, W. *J. Bact.* **171**, 2142–2147 (1989).
- McRee, D. E. *J. molec. Graph.* **10**, 44–47 (1992).
- Furey, W. & Swaminathan, S. in *ACA Meeting Abstr.* 73 (Am. Crystallogr. Assoc., New York, 1990).
- Read, R. J. *Acta crystallogr. A* **42**, 140–149 (1986).
- Brünger, A. T. *X-PLOR: A System for X-ray Crystallography and NMR* (Yale Univ. Press., New Haven, 1992).
- Segal, E., Billyard, E., So, M., Storzbach, S. & Meyer, T. F. *Cell* **40**, 293–300 (1985).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J. appl. Crystallogr.* **26**, 283–291 (1993).
- Carson, M. J. *appl. Crystallogr.* **24**, 958 (1991).
- Swanson, J. & Koomey, J. M. in *Mobile DNA* (eds Berg, D. & Howe, M.) 743–761 (Am. Soc. Microbiol., Washington DC, 1989).
- Connolly, M. L. *Science* **221**, 709–713 (1983).
- Gilson, M. K. & Honig, B. *Proteins: Struct. Funct. Gen.* **4**, 7–18 (1988).

ACKNOWLEDGEMENTS. We thank the staff at SSRL for help with synchrotron data collection; M. So and C. D. Deal for discussion; S. L. Bernstein, M. A. Capozza, B. W. Kays and K. A. Slater for pilin production and purification; J. J. Lavenue, B. R. Crane, M. A. McTigue, A. S. Arvai and D. E. McRee for SSRL data collection support; Y. Bourne for aiding the disaccharide interpretation; T. D. Hong for film scanning; M. E. Pique and T. J. Macke for graphics support; and C. L. Fisher for help with DISCOVER calculations. This research was supported by an NIH grant to J.A.T. and an NIH postdoctoral fellowship to K.T.F.

An ultraviolet flare at the centre of the elliptical galaxy NGC4552

Alvio Renzini^{*†}, Laura Greggio[†],
Sperello di Serego Alighieri[‡], Michele Cappellari[§],
David Burstein[¶] & Francesco Bertola[§]

^{*} European Southern Observatory, D-85748 Garching bei München, Germany

[†] Department of Astronomy, University of Bologna, I-40126 Bologna, Italy

[‡] Osservatorio Astrofisico di Arcetri, Firenze, Italy

[§] Department of Astronomy, University of Padova, Italy

[¶] Department of Physics and Astronomy, Arizona State University, Tempe, Arizona, USA

Most present-day galaxies are not 'active', in that they show no signs of continuing, high-energy events. The high energy output of an active galaxy is usually attributed to accretion of gas onto a massive black hole at its centre. Yet it remains an open question whether such black holes might exist at the centre of most large galaxies, and be undetected because no gas is presently being accreted to power a nuclear emission¹. Here we report the detection of an ultraviolet flare at the centre of NGC4552, an otherwise optically quiescent elliptical galaxy. The flare reached a luminosity about one million times that of the Sun, and probably arose from a mild accretion event onto a central black hole. The accreted gas could have come either from a star passing nearby, or from an interstellar cloud. This serendipitous discovery suggests that ultraviolet flares near the centres of galaxies may be common events,

and offers a new way to search for black-hole-related activities in otherwise quiescent galaxies.

We have imaged the elliptical galaxies NGC1399 and NGC4552 with the pre-COSTAR Hubble Space Telescope Faint Object Camera (FOC) in the F96 zoomed mode, with the aim of studying the origin of the ultraviolet light in these systems². These two galaxies are generally viewed as being very similar in the optical region of the spectrum and relatively quiescent in terms of their stellar populations². Images for both galaxies were obtained through four FOC filters: F175W (peak wavelength $\lambda = 1,720 \text{ \AA}$); F220W (2,270 $\text{\AA}$); F275W (2,760 $\text{\AA}$); and F342W (3,400 $\text{\AA}$). All four images of NGC4552 contain prominent central spikes that appear point-like (that is, unresolved by the ~ 0.07 -arcsec core of the aberrated Hubble Space Telescope (HST) point-spread function). No such spike is present in the four images of NGC1399. The analysis of the full data set will be presented elsewhere (F.B. *et al.*, manuscript in preparation); here we report the unexpected finding that the ultraviolet-bright central spike in NGC4552 is actually a flare.

Figure 1 shows the central parts of the FOC-F342W images of NGC4552 taken by us in 1993 (left) and by Crane *et al.*³ in 1991 (right). The ultraviolet spike in the 1993 image is detected with a signal-to-noise ratio of about 11, and appears to coincide with the centroid of the isophotes within one pixel (0.02 arcsec, or $\lesssim 1.6 \text{ pc}$ adopting a distance $D = 16.8 \text{ Mpc}$). In contrast, the 1991 image shows two fainter, nearly equal peaks, each with a signal-to-noise ratio of ~ 4 above the local background, separated by $\sim 0.14 \text{ arcsec}$ ($\sim 10 \text{ pc}$). The upper peak in Fig. 1 (right) coincides within one pixel of the centroid of the isophotes (and therefore with the 1993 spike). The upper and lower peaks in the Crane *et al.* image are here referred to as the central and off-centre peak, respectively. Crane *et al.* had noticed these two

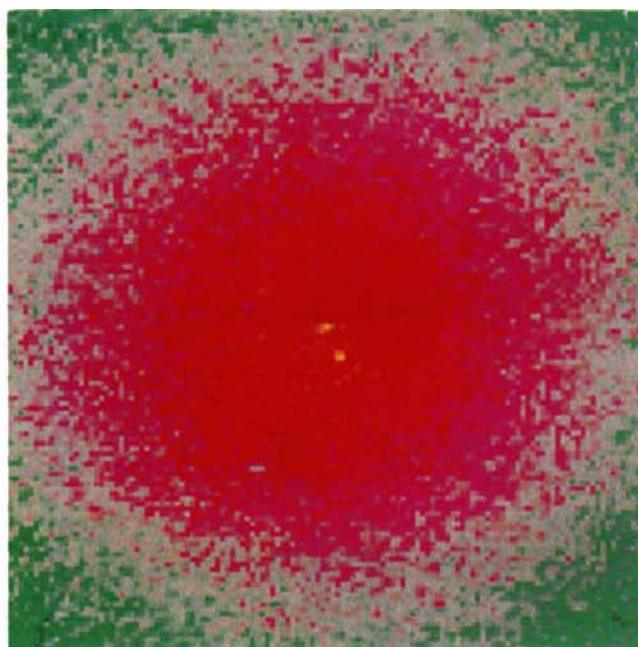
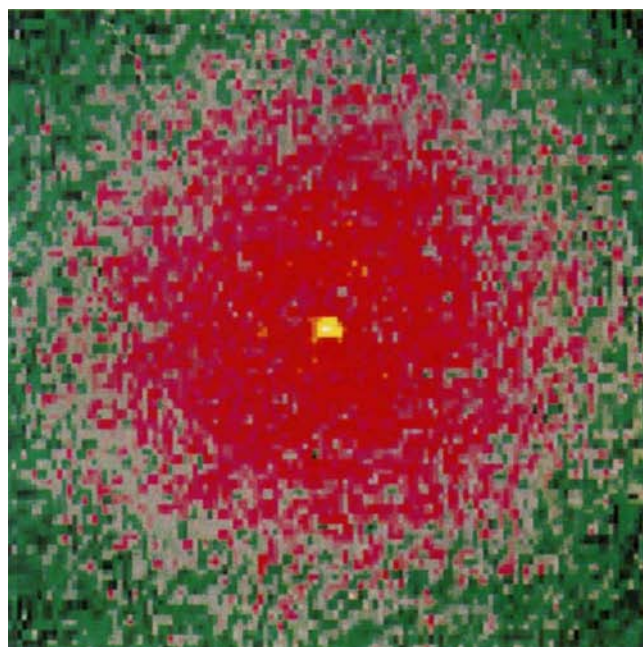


FIG. 1 The FOC/F96/F342W (U-band) images of the central regions of NGC4552 taken by us on 28 November 1993 (left), and by Crane *et al.*³ on 19 July 1991 (right). The 1993 image was obtained with the FOC in the $512 \times 1,024$ zoomed mode, with original pixel size of $0.044 \times 0.022 \text{ arcsec}$ and with an exposure time of 596.6 s. The 1991 image was obtained in 512×512 unzoomed mode, with pixel size of $0.022 \times 0.022 \text{ arcsec}$ and an exposure time of 1,196 s. The 1993 image shown here had been rebinned (de-zoomed) so to have the same pixel size as the 1991 image. Besides blemishes and reseau mark removal no other image processing has been applied (specifically, these images are not deconvolved). Each panel is 2.9 arcsec on a side, north is up and east at the left. The central peak gave a total of 145 ± 32

counts above the background in 1991, and 350 ± 31 counts in 1993 (with about 50% of the exposure time). These counts refer to circular apertures of 1.5 and 2 pixels radius, so as to allow for the different point-spread function respectively in the unzoomed (1991) and de-zoomed (1993) images. (With a 2-pixel aperture, the 1991 central spike would have given 222 ± 56 counts.) All errors include the uncertainty in estimating the background. The "jitter plot" for the 1991 exposure shows that the HST pointing was within one pixel during the whole exposure—actually, within $\sim 1/5$ of a pixel over 90% of the exposure. This ensures that a genuine variation is the only explanation for the brightening in the central peak between 1991 and 1993.

peaks in their analysis, but were unsure of their reality (M. Stiavelli, personal communication). A visual-filter (F502M) image taken in 1991 at the same time as the F342W image³ does not show evidence of either peak, implying that they were bluer than $U-V \approx 0.5$ (1σ -level). Our 1993 images recover only the central 1991 peak, even though the two 1991 peaks have identical statistical significance.

Figure 2 shows the surface brightness profiles of the central regions of NGC4552 for the 1991 and 1993 F342W images of Fig. 1. A quantitative comparison of the two images indicates that the central peak has brightened by a factor of 7 ± 1.5 over this ~ 2 -year period. The actual variation is probably somewhat larger than this, as we are conservative in correcting for non-linearity effects⁴ in the 1993 image. We also note that if the off-centre peak were still of the same brightness as in 1991, it would have been detected only at the $\sim 2.5\sigma$ level in the 1993 image. Based on current evidence, the off-centre peak seen in 1991 yields little further insight into the source of the ultraviolet central spike.

The count rates through the four filters, coupled with convolutions of either a power law flux $F_\lambda \propto \lambda^\alpha$ or a black-body spectrum

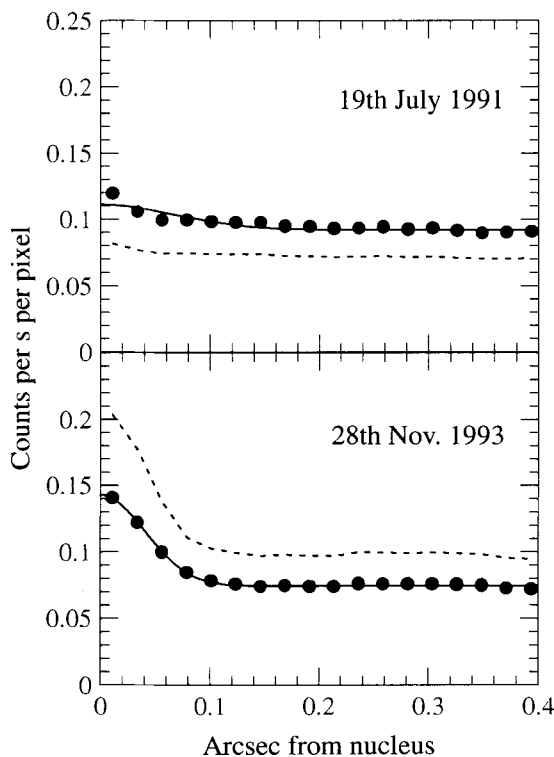


FIG. 2 The isophotal surface brightness profiles for the innermost 0.4 arcsec of the 1991 image (top) and the 1993 image (bottom). The dashed line in the bottom panel represents the count rate after correction for nonlinearity effects. The dashed line in the upper panel shows the expected count rate, had the 1991 observation been made with the same zoomed mode used for the 1993 image. For both images the centre has been taken at the pixel of maximum count rate, that for the 1991 exposure coincides with the central peak, and is within one pixel of the isophotal centre. The dot-dashed line in the bottom panel shows a 0.07-arcsec FWHM gaussian which approximates the central core of the spherically aberrated point-spread function of the HST FOC. The central spike appears essentially unresolved, given the saturation effects and the original pixellation in the zoomed mode. Applying the appropriate nonlinearity corrections⁴ we estimate that the 1993 spike count rate should be increased by at least a factor ~ 1.8 ; so it would have given $(350 \pm 31) \times 2 \times 1.8/1.25 = 1,008 \pm 89$ counts, if observed in the same mode and same integration time used for the 1991 image (the FOC sensitivity in the zoomed mode is a factor of 1.25 higher than in the unzoomed mode).

with the filter response functions, allow estimates of the ultraviolet spectral energy distribution of the 1993 spike. We get equally good fits for either $\alpha = 0.0 \pm 0.4$ or $T = 12,000 \pm 1,000$ K, with the flux in the 1,750–3,400 Å range being nearly flat at the level of $\sim 2.2 \times 10^{-17}$ erg cm⁻² s⁻¹ Å⁻¹. The black-body distribution yields a bolometric luminosity $\sim 10^6 L_\odot$ (where $L_\odot$ is the solar luminosity).

We briefly evaluate four candidate interpretations of the flare in NGC4552: (1) a nova or supernova outburst, (2) a stellar collision, (3) a gravitational lensing event, or (4) a transient accretion event onto a central massive black hole.

(1) Novae are typically 10^4 – $10^5 L_\odot$ at maximum light, too faint to produce the central spike ($>10^6 L_\odot$). Although supernovae are much brighter, based on the observed supernova rate in ellipticals⁵ and the NGC4552 central surface brightness³, one supernova is expected to explode every $\sim 3 \times 10^7$ yr within 0.02 arcsec ($\lesssim 1.6$ pc) from the centre of this galaxy. Moreover, the ultraviolet colours and luminosity of the 1993 spike do not match those of type Ia supernovae⁶, the only ones ever observed in elliptical galaxies.

(2) The luminosity (and hence the stellar) density at the centre of NGC4552 is rather low³, some 10 times lower than in high-density globular clusters, where stellar collisions are already predicted⁷ to be as rare as one event every 10^8 – 10^9 yr.

(3) Gravitational lensing is an attractive option, as it might explain both the two 1991 peaks, and the 1993 brightening. However, the two peaks are not symmetric with respect to the centre of the galaxy, and the peak that has brightened has not changed position. A very contrived lensing scenario would likely be required. In addition, amplification by a factor of $\sim 10^3$ – 10^4 (hence an extremely close alignment) is needed to produce an apparent $10^6 L_\odot$ image, if the lensed object is a $\sim 10^4$ K (a few times $10^2 L_\odot$) main-sequence star. The probability of such an event is vanishingly low, even assuming that a spiral galaxy in the Virgo cluster is right behind NGC4552 (and there is no evidence for such a situation existing). Finally, the lensing of a distant quasar is another event of extremely low probability⁸.

(4) The occurrence of central, ultraviolet-bright flares has been predicted to be a rather common event, the result of the tidal disruption—or stripping—of a star in a close fly-by with the black hole^{1,9}, provided (of course) that most spheroids host such objects. One total disruption every $\sim 10^3$ – 10^4 yr has been estimated for the central black hole in a giant elliptical galaxy^{1,9} and partial stripping events should be more frequent, resulting from looser star/black-hole fly-bys. If these estimates are correct, then an event of this kind is not only far more frequent than any of the other phenomena mentioned above, but would also match the ultraviolet-bright spectral energy distribution of the observed flare^{1,9}. To date, only the limiting case of the total disruption of a main-sequence star has been investigated. The resulting flare is very bright for several years (a few times $10^{10} M_6 L_\odot$, where M_6 is the black-hole mass in units of $10^6 M_\odot$)^{1,9–12}. The fact that in 1993 the flare was much fainter requires that the fly-by was rather wide, that the star did not fully disrupt, and that only $\sim 10^{-3} M_\odot$ was stripped (the luminosity scales as the 5/3 power of the stripped mass¹⁰).

An alternative black-hole-accretion scenario, in which the mass comes from the interstellar medium, finds circumstantial support from the existence of extended, centrally peaked H α emission¹³ in the inner ~ 2 kpc of NGC4552. Thus, this stream of gas clouds (the likely result of a dwarf galaxy having been recently captured) may occasionally fuel a central black hole. As is common for many galaxies, NGC4552 hosts a very weak, compact radio source variable on timescales of several years^{14,15} that might also be variable at 10 μ m (ref 16). Thus while some weak nuclear activity existed in NGC4552 even before the 1993 flare, NGC1399 is a much stronger radio source¹⁷ yet the HST images reveal no central spike.

Whatever the origin of the NGC4552 flare, it is a case of nuclear activity at an unprecedentedly faint level, that the only

HST could have revealed thanks to its unique combination of angular resolution and sensitivity to ultraviolet. The fact that we know the central ultraviolet-spike in NGC4552 is a flare owes more to chance than to any scientific planning, as this may well be the only normal galaxy ever imaged at two widely separated epochs in the ultraviolet by HST.

With few exceptions^{18,19} direct evidence that bright galaxies house central massive black holes is still elusive. The refurbished HST can now detect ultraviolet-bright central spikes much fainter than that found in NGC4552. Ultraviolet imaging by the HST therefore offers the opportunity for a systematic survey of this new kind of activity at the centres of galaxies, and another chance to discover whether most giant galaxies harbour central massive black holes. □

Received 17 August; accepted 26 September 1995.

1. Rees, M. *Nature* **333**, 523–528 (1988).
2. Burstein, D., Bertola, F., Buson, L. M., Faber, S. M. & Lauer, T. R. *Astrophys. J.* **328**, 440–462 (1988).
3. Crane, P. et al. *Astr. J.* **106**, 1371–1393 (1993).
4. Greenfield, P. *Instrument Science Report FOC-074* (Space Telescope Science Institute, Baltimore, MD, 1994).
5. Turatto, M., Cappellaro, E. & Benetti, S. *Astr. J.* **108**, 202–206 (1994).
6. Panagia, N. & Gilmozzi, R. in *Supernovae* (ed. Woosley, S. E.) 497–502 (Springer, New York, 1992).
7. Hills, J. G. & Day, C. A. *Astrophys. Lett.* **17**, 87–93 (1976).
8. Schneider, D. P. et al. *Astr. J.* **95**, 1619–1628 (1988).
9. Rees, M. *Science* **247**, 817–823 (1990).
10. Canizzo, J. K., Lee, H. M. & Goodman, J. *Astrophys. J.* **351**, 38–46 (1990).
11. Phinney, E. S. in *The Center of the Galaxy* (ed. Morris, M.) 543–553 (Kluwer, Dordrecht, 1989).
12. Kochanek, C. S. *Astrophys. J.* **422**, 508–520 (1994).
13. Trinchieri, G. & di Serego Alighieri, S. *Astr. J.* **101**, 1647–1658 (1991).
14. Jenkins, C. R. *Mon. Not. R. astr. Soc.* **200**, 705–731 (1982).
15. Wrobel, J. B. *Astr. J.* **101**, 127–147 (1991).
16. Sparks, W. B., Hough, J. H., Axon, D. J. & Bailey, J. *Mon. Not. R. astr. Soc.* **218**, 429–444 (1986).
17. Bolton, J. G., Wright, A. E. & Savage, A. *Austr. J. Phys., Astrophys. Suppl.* **46**, 1–17 (1979).
18. Ford, H. C. et al. *Astrophys. J.* **435**, L27–L36 (1994).
19. Miyoshi, M. et al. *Nature* **373**, 127–129 (1995).

ACKNOWLEDGEMENTS. We thank A. Caulet for initial help in data reduction, and P. Greenfield, W. Hack, A. Nota and R. Jedrzejewski of STScI for their help in gauging the FOC nonlinearity effects. We also thank T. Broadhurst, P. Parma, M. Rees, M. Stiavelli and R. Woistencroft for useful discussions. This work was supported by the Italian Ministry of University (MURST), the Italian Space Agency (ASI), and by NASA (to D.B.).

Diamond and silicon carbide in impact melt rock from the Ries impact crater

R. M. Hough*, I. Gilmour*†, C. T. Pillinger*, J. W. Arden†, K. W. R. Gilkes§, J. Yuan|| & H. J. Milledge¶

* Planetary Sciences Unit, Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

† Department of Earth Sciences, Oxford University, Parks Road, Oxford OX1 3PR, UK

§ School of Physics, University of East Anglia, Norwich NR4 7TJ, UK

|| Cavendish Laboratory, Cambridge University, Cambridge CB3 0HE, UK

¶ Department of Geological Sciences, University College, Gower Street, London WC1E 6BT, UK

SHOCK-PRODUCED diamond and lonsdaleite (the hexagonal polymorph) were first observed in experiments involving explosions¹. Several classes of meteorites^{2,3} contain microcrystalline diamond aggregates that are thought to be produced by impacts with the Earth or in space. Diamonds have also been found in association with several Russian impact craters⁴ and in Cretaceous/Tertiary boundary impact ejecta^{5,6}; these too have most often been interpreted as having formed by shock in the solid state⁴. Here we report

the occurrence of diamond/lonsdaleite plates and cubic diamond in association with silicon carbide, in impact melts from the Ries crater in southern Germany. We interpret these occurrences as evidence that these phases can be formed by chemical vapour deposition from the ejecta plume of an impact crater. It follows that cubic diamond and silicon carbide may be formed at any impact site from vaporized carbon-bearing rocks, and hence may be used as a reliable diagnostic tool for hypervelocity impact on Earth. This process may also explain the occurrence of diamonds found in sediments (carbonados⁷), which may result from the 'heavy bombardment' period of early Earth history, rather than from mantle-derived diatremes⁸.

An early Russian report⁹ on the existence of gas inclusions in glass from the Ries crater mentions finding grains containing high-pressure carbon polymorphs, one of which was shown to be diamond/lonsdaleite by X-ray diffraction. Recently, Masaitis⁴ reported finding a small number of apographitic diamonds in the impact melt rock of the Ries crater, and similar diamonds have been found at the Popigai crater in Russia¹⁰. Previous careful studies of the shocked graphite-bearing gneisses from the Ries in the 1960s (A. El Goresy, D. Stöffler, personal communication) had proven negative, except for the identification of chaoite¹¹, controversially claimed to be the *sp*-bonded allotrope of carbon. As acid dissolution studies have been extremely successful in isolating robust carbon phases, particularly tiny diamonds, from primitive meteorites^{12,13} and from the iridium-rich layers at the Cretaceous/Tertiary boundary^{6,14}, we investigated their applicability to recovering diamond polymorphs from the Ries ejecta.

The Ries crater (48° 51' 06" N, 10° 29' 23" E) is a 24-km-diameter impact feature in southern Germany which formed 15 Myr ago. We collected pieces of fallout impact melt rocks, called suevite, from the Otting quarry (3.5 km from the crater rim). The rock fragments as recovered were pre-etched in 3 M HCl and dilute HF/HNO₃ to remove 10% of the total mass as a precaution against various contaminating artefacts. Lumps (335 g) were then broken up and treated cyclically with HF, HCl, Cr₂O₇²⁻, HClO₄ and H₂SO₄ under conditions designed to destroy silicates, organics, graphite, rutile and zircon. The carbon-isotope composition was monitored, and found to be ~ -22 to -24‰ (relative to the PDB standard¹⁵) at various stages; the carbon content of the residue gradually increased to 80 wt%. A stepped combustion¹⁶ performed before final removal of zircon afforded a sharp release of carbon, equivalent to 19.1 wt% of the sample, over the temperature range 500–700 °C (Fig. 1), but the isotopic composition of the gas released suggested two

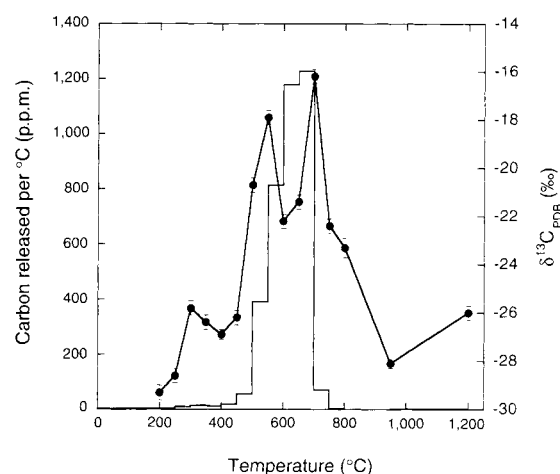


FIG. 1 Stepped combustion profile of carbon from the Ries Otting impact melt rock (suevite) acid-resistant residue before removal of zircon. The histogram gives carbon yield information, and the line graph with filled circles corresponds to the isotopic composition at each temperature step.

† To whom correspondence should be addressed.

carbon phases different in ^{13}C abundance by possibly 6‰ (–16‰ versus –22‰). There was little or no carbon burning at temperatures greater than 750 °C. The expected combustion temperature of micrometre-sized SiC is well over 1,000 °C (ref. 17), implying that SiC and diamond co-combust, possibly because of the fine-grained nature of the SiC or as a result of the exothermic nature of diamond burning.

A binocular-microscope examination of the residue showed it to contain transparent plate-like grains with a brownish to yellow tinge (grain A in Fig. 2), green-to-blue crystals (grain B in Fig. 2), both up to 100 µm in size, and a majority of smaller dark cinder-like and weakly coherent fragments from a few micrometres up to a maximum of 200 µm in size (composite grains C in Fig. 2). Single-grain X-ray analysis showed that the plate-like grains were mainly diamond with X-ray diffraction patterns that gave reasonable {100} and diffuse {101} lonsdaleite reflections¹⁸. The blue-greenish crystals were very well crystallized and comprised hexagonal α -SiC; one that we studied in detail was entirely of the 4H polytype. The composite grains were a mixture of diamond and SiC. The SiC was 4H type and relatively coarse grained; those lines in the powder pattern which correspond to the cubic (β -SiC) phase were enhanced, suggesting the presence of fine-grained β -SiC that has not yet been found among the single crystals. Because of the ubiquity of SiC it was difficult to be sure whether lonsdaleite occurred in the fine-grained specimens.

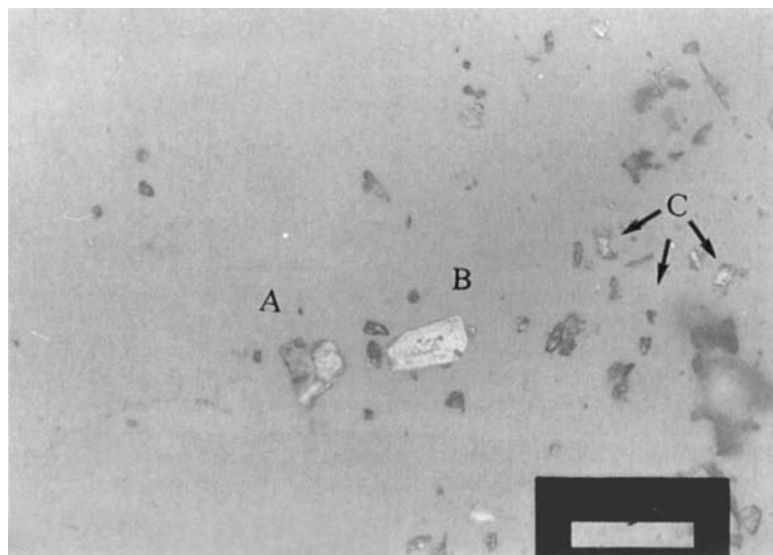
The residue was also investigated by analytical transmission electron microscopy (TEM). The major component of finer material was composed of skeletal aggregates, typically a few micrometres in their largest dimensions. They contained diamonds ranging in size from tens of nanometres to 2 µm (Fig. 3a) and variably sized SiC, but no lonsdaleite. The larger diamonds were often in the form of platelets with {111} orientation; electron energy-loss spectroscopy (EELS) revealed fine structure following the carbon K ionization edge at 290 eV, confirming that the platelets are pure cubic diamond¹⁹. Single crystals of probably 4H α -SiC, up to 1 µm in size, were also observed in the TEM (Fig. 3b). Diamond and SiC crystallites, identified by selected-area electron diffraction (SAED) and EELS, were often observed in the same skeletal aggregate, and appeared to be either overlapping or intergrown epitaxially (Fig. 3c). Silicon was the only element other than carbon seen during analysis of grains in the electron microscope.

The above findings confirm the existence of diamond/lonsdaleite at the Ries^{4,9} crater, similar to the known occurrences of diamond/lonsdaleite in impact melt rocks from the Popigai

crater^{10,20} where they are believed to result from the direct shock conversion of graphite to diamond/lonsdaleite. However, quite distinct from all previous studies concerning diamond of impact origin, we find that the vast majority (~3 p.p.m.) of the mineral is in skeletal aggregates and for the first time observe a close association in these grains between diamond and silicon carbide. The presence of SiC in the aggregates and as single grains is unlikely to be a product of shock. It is not mentioned amongst the products of shock experiments performed under a variety of conditions to produce diamond²¹. The known history of our sample, the precautions taken to avoid artefacts and previous work on identifying synthetic carborundum powder by stepped combustion¹⁷ makes it unlikely that the SiC is a contaminant.

It is our contention that both the diamond and the silicon carbide are indigenous to the sample, and as such must have been made by a process associated with the impact. In this respect we favour a vapour deposition mechanism. In chemical vapour deposition (CVD) experiments, it is well known that diamond grows on silicon-containing substrates²² and that it does so via an interfacial layer of 4H-SiC (J. Hutchison, personal communication). Hexagonal α -SiC has also been shown to grow on diamond in high-temperature experiments²³. β -SiC and 2H-SiC are believed to be the low-temperature forms of the mineral (produced from 1,500 to 2,000 °C) whereas α -SiC (4H, 6H, 8H) are the higher-temperature polytypes²⁴. We note that 4H-SiC apparently has the highest growth rate of any of the silicon carbide polytypes during epitaxial growth by sublimation on to Si surfaces²⁵. The skeletal form of the Ries diamond suggests that it grew free or between the surfaces of other minerals. A stepped combustion of an HF/HCl-resistant residue gave a much broader temperature range for the carbon release than was obtained for the final demineralized sample (Fig. 1), implying intergrowth of carbon- and non-carbon-containing species. The impact of a kilometre-sized body, at the average velocity of Earth impactors of 25 km s⁻¹, into sedimentary strata overlying crystalline graphitic gneiss basement would undoubtedly produce a vapour, or even a plasma, containing H, C and Si, in elementary or ionized form, exactly the feed stock for diamond and SiC formation by CVD. The isotopic composition of the diamond/SiC is in keeping with what might be expected from mixing of carbonate (~0‰) sources derived from the 600 m of Mesozoic sediments in the Ries region at the time of the impact and an organic/graphitic (~–28‰) source derived from the sediments, underlying gneisses or the impactor itself. The extent of any isotopic fractionation associated with the vaporization of these source materials is unknown. Taking into consideration

FIG. 2 Types of grains located by binocular examination of the final acid-resistant residue of Ries crater suevite from Otting. A, platey diamonds with minor amounts of lonsdaleite; B, 4H-SiC; and C, friable aggregates of diamond plus silicon carbide, probably both 4H and β type. Scale bar (clear bar inside black rectangle), 170 µm.



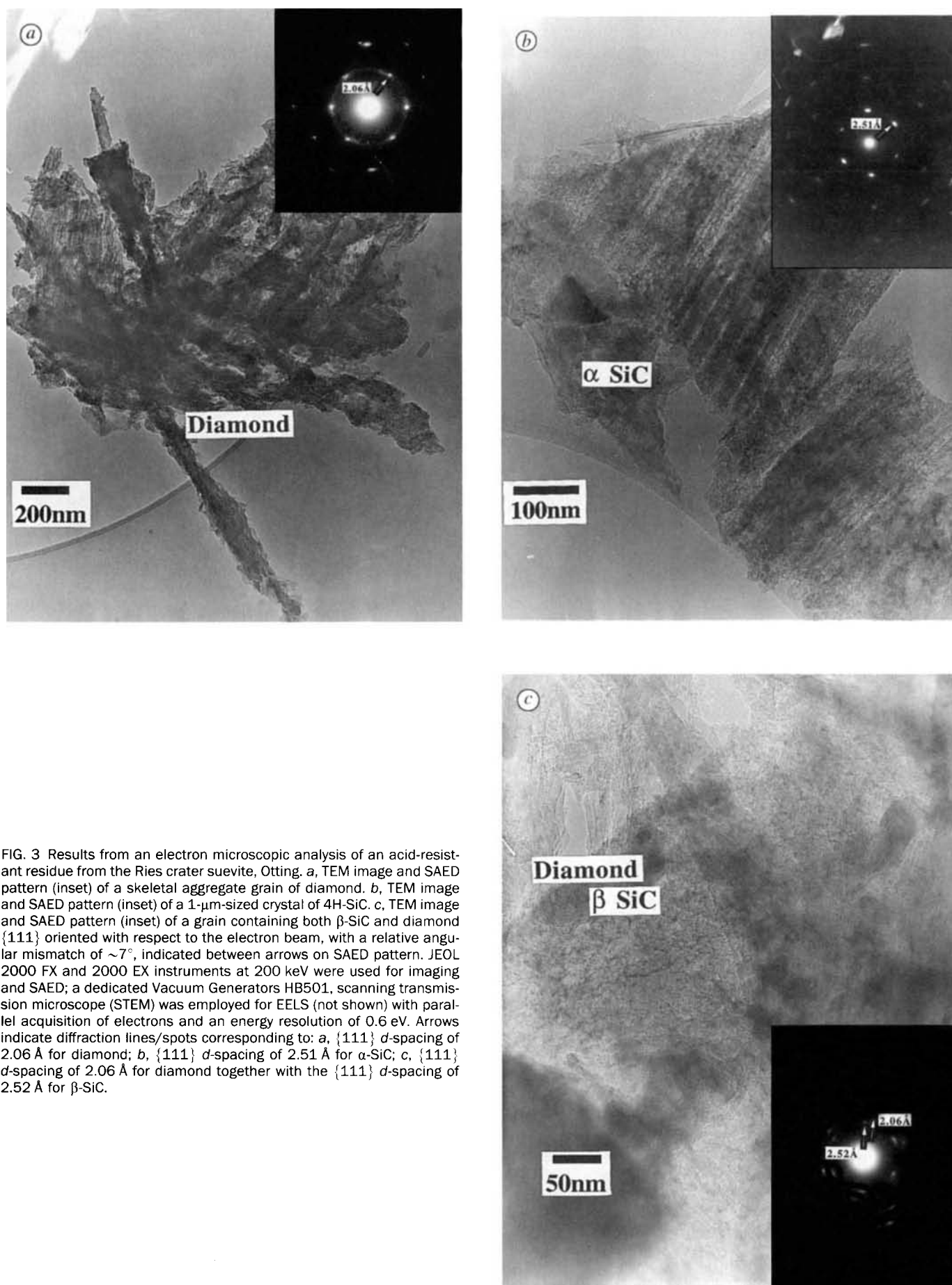


FIG. 3 Results from an electron microscopic analysis of an acid-resistant residue from the Ries crater suevite, Otting. *a*, TEM image and SAED pattern (inset) of a skeletal aggregate grain of diamond. *b*, TEM image and SAED pattern (inset) of a $1\text{-}\mu\text{m}$ -sized crystal of 4H-SiC . *c*, TEM image and SAED pattern (inset) of a grain containing both $\beta\text{-SiC}$ and diamond $\{111\}$ oriented with respect to the electron beam, with a relative angular mismatch of $\sim 7^\circ$, indicated between arrows on SAED pattern. JEOL 2000 FX and 2000 EX instruments at 200 keV were used for imaging and SAED; a dedicated Vacuum Generators HB501, scanning transmission microscope (STEM) was employed for EELS (not shown) with parallel acquisition of electrons and an energy resolution of 0.6 eV. Arrows indicate diffraction lines/spots corresponding to: *a*, $\{111\}$ d -spacing of $2.06\text{ \AA}$ for diamond; *b*, $\{111\}$ d -spacing of $2.51\text{ \AA}$ for $\alpha\text{-SiC}$; *c*, $\{111\}$ d -spacing of $2.06\text{ \AA}$ for diamond together with the $\{111\}$ d -spacing of $2.52\text{ \AA}$ for $\beta\text{-SiC}$.

just the crater suevite at the Ries, we calculate 7.2×10^4 tonnes of diamond/SiC in 3:1 proportions exists in the area. Ample carbon is available in either the basement or the projectile; each would need to contain only 6.5 p.p.m. C to provide the source material although additional C may be required to produce sufficiently reducing conditions for SiC formation. Acquiring enough carbon from the sediment would be even simpler, but the $\delta^{13}\text{C}$ of the acid residue argues against the majority coming from carbonate. Masaitis²⁰ has argued that graphite- (or coal-) bearing targets are a prerequisite for diamond formation during impacts. There is no evidence at the Ries for sedimentary fragments shocked to pressures greater than 10 GPa, although the

basement gneisses undoubtedly were²⁶, and these may be the source of the plate-like diamond/lonsdaleite crystals, as the suevite that we dissolved contained basement inclusions.

Given the ubiquity of the diamond polymorphs at several Russian craters, the Ries crater and the K/T boundary, perhaps the presence of diamond in its broadest sense (and possibly SiC) should become criteria for the identification of impact structures, as are coesite, stishovite, planar deformation features in quartz and shatter cones. The value of diamonds in this respect is their ability to survive over immense periods of geological time, giving access to events that occurred (and possibly carbon reservoirs that existed) in the early history of the Earth when bombardment was as its most intense. □

Received 4 April; accepted 25 September 1995.

- DeCarli, P. S. & Jamieson, J. C. *Science* **133**, 1821–1822 (1961).
- Lipschutz, M. E. & Anders, E. *Geochim. cosmochim. Acta* **24**, 83–105 (1961).
- Vdovykin, G. P. *Space Sci. Rev.* **10**, 483–510 (1970).
- Masaitis, V. L. *Meteoritics* **27**, 21–27 (1992).
- Carlisle, D. B. & Braman, D. R. *Nature* **352**, 708–709 (1991).
- Gilmour, I. et al. *Science* **258**, 1624–1626 (1992).
- Ozima, M., Zashu, S., Tomura, K. & Matsuhisa, Y. *Nature* **351**, 472–474 (1991).
- Smith, J. V. & Dawson, J. B. *Geology* **13**, 342–343 (1985).
- Rost, R., Dolgov, Y. A. & Vishnevskiy, S. A. *Dokl. Acad. Nauk. USSR* **241**, 165–168 (1978).
- Koeberl, C. et al. *Lunar planet. Sci.* **26**, 777–778 (1995).
- El Goresy, A. & Donnay, G. *Science* **161**, 363–364 (1968).
- Pillinger, C. T. *Phil. Trans. R. Soc. Lond. A* **343**, 73–86 (1993).
- Anders, E. & Zinner, E. *Meteoritics* **28**, 490–514 (1993).
- Hough, R. M., Wright, I. P., Pillinger, C. T. & Gilmour, I. *Lunar planet. Sci.* **26**, 629–630 (1995).
- Craig, H. *Geochim. cosmochim. Acta* **12**, 133–149 (1957).
- Swart, P. K., Grady, M. M. & Pillinger, C. T. *Meteoritics* **18**, 137–154 (1983).

- Russell, S. S. & Pillinger, C. T. *J. chem. Soc., Faraday Trans.* **89**, 2297–2304 (1993).
- Yagi, T., Utsumi, W., Yamakata, M., Kikegawa, T. & Shimomura, O. *Phys. Rev.* **B46**, 6031–6039 (1992).
- Berger, S. D., McKenzie, D. R. & Martin, P. J. *Phil. Mag. Lett.* **57**, 285–290 (1988).
- Masaitis, V. L. *Regionalnaia Geologia i Metallogenia* **1**, 121–134 (1993).
- Heimann, R. B. & Kleiman, J. in *Crystals: Growth, Properties and Applications* (ed. Freyhardt, H. C.) 1–76 (Springer, Berlin, 1988).
- Sato, Y. & Kamo, M. in *The Properties of Natural and Synthetic Diamond* (ed. Field, J. E.) 423–469 (Academic, London, 1992).
- Varshavskii, A. V. & Shul'pyakov, Y. F. *Soviet Phys. Dokl.* **12**, 202–204 (1967).
- Yamada, K. & Tobisawa, S. *Phil. Mag.* **A61**, 943–954 (1990).
- Kalnin, A. A., Neubert, F. & Pezoldt, J. *Diamond Related Mater.* **3**, 346–352 (1994).
- von Engelhardt, W., Luft, E., Arndt, J., Schock, H. & Weiskirchner, W. *Geochim. cosmochim. Acta* **51**, 1425–1443 (1987).

ACKNOWLEDGEMENTS. We thank N. Williams for help with electron microscopy and S. O. Agrell for advice concerning Ries samples. We are grateful to D. Stöffler for a very constructive review that significantly improved this manuscript. This work was supported by PPARC and by a Royal Society University Research Fellowship (I.G.). R.M.H. thanks NERC for a studentship.

Stability of O_2/H_2 mixtures at high pressure

Paul Loubeyre & René LeToullec

Physique des Milieux Condensés (CNRS URA 782), boîte 77, Université Paris 6, 4 place Jussieu, 75252 Paris, France

THE discovery^{1–3} of the stoichiometric compounds $\text{He}(\text{N}_2)_{11}$, $\text{Ne}(\text{He})_2$ and $\text{Ar}(\text{H}_2)_2$ in high-pressure mixtures has uncovered a new aspect of high-pressure chemistry. In contrast to these studies on inert species, we show here that high pressures can lead to unexpected behaviour in reactive compounds. We find that mixtures of O_2 and H_2 are stable at pressures of around 7.6 GPa, despite their explosive nature under ambient conditions. We use microscopy and Raman spectroscopy to identify the phase boundary of the stable region. The structure of the binary phase diagram at 296 K suggests that a stoichiometric compound $(\text{O}_2)_3(\text{H}_2)_4$ might exist. The stability of dense O_2/H_2 mixtures might be usefully exploited in the development of rocket fuels and for energy storage, and might also be relevant to the composition of the interiors of the jovian planets.

The properties of molecular hydrogen and oxygen have been much investigated at high pressures in the diamond-anvil cell^{4,5}. But the explosive nature of H_2/O_2 mixtures at ambient pressure has fostered the perception that the high-pressure properties of this binary system are likely to hold little of interest. Our finding that, at high pressure, such mixtures do not necessarily react to form water but instead may show more rich and subtle behaviour therefore comes as a surprise.

We have used a membrane diamond-anvil cell that is well adapted for probing phase transitions at fixed temperature⁶. The cell was loaded in a pressure vessel at room temperature but, for safety reasons, under pressures of around just 90 bar, a factor of ten smaller than the pressures generally used for this technique. (During one loading, probably with too fast a pressure increase, an explosion occurred in the pressure vessel and caused major damage.) Concentrations were calculated from the partial pressures in the initial mixture, corrected using the second virial

coefficients. BeCu gaskets were used for the sample chamber, which was typically 50 μm in diameter and 15 μm thick at 10 GPa. The pressure was measured by the quasi-hydrostatic ruby scale⁷.

When loaded in the diamond-anvil cell, the O_2/H_2 mixtures were stable, or reacted only very slowly, in the dense fluid and the solid phase over a period of a month. A 488-nm Ar^+ laser, a halogen lamp and even a synchrotron X-ray beam (if applied for less than a few hours) were all unable to initiate a reaction. Raman spectra from 1,400 to 5,000 cm^{-1} showed no vibron peak characteristic of reaction products such as H_2O or H_2O_2 . (When slight reaction could be induced in the fluid phase, well below the solidification pressure, by focusing the laser on the edge of the sample chamber, phase separation and the stretching mode of H_2O could be observed.)

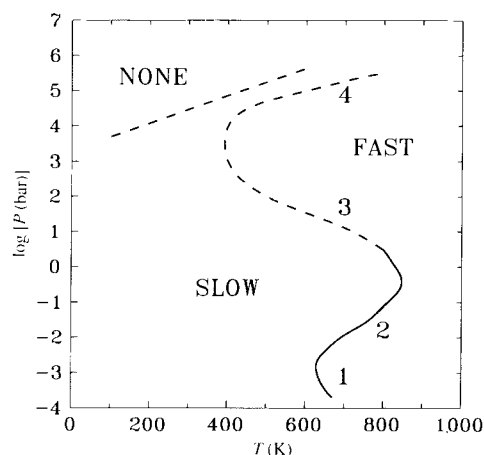


FIG. 1 Explosion limits of a stoichiometric H_2/O_2 mixture. The solid line represents available experimental data⁸ on the boundary between slow and fast reaction. The dashed lines indicate that the boundaries are uncertain but have a shape inferred from the present study. We suggest the existence of a fourth explosion limit, and a boundary between slow reactions and no reaction.

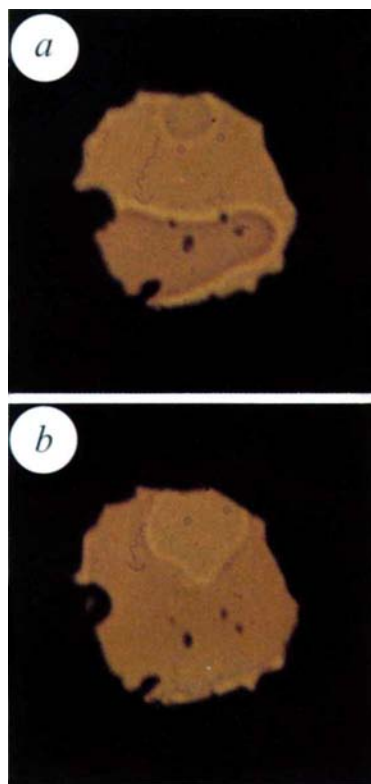


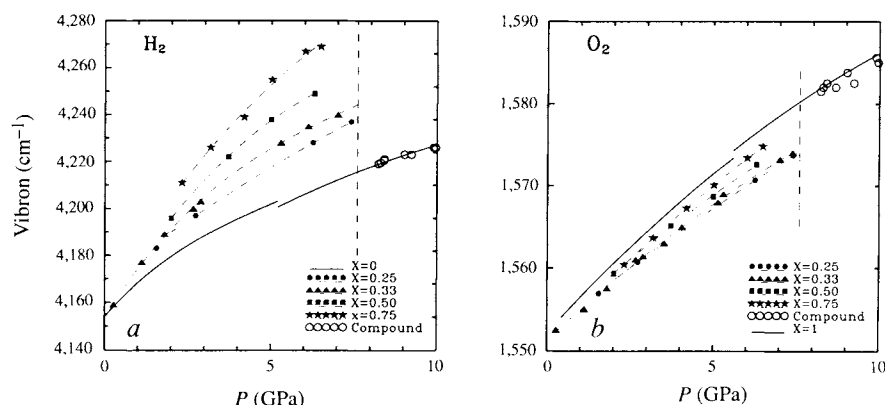
FIG. 2 Photomicrographs of the growth of a single crystal of O_2/H_2 alloy in equilibrium with a small almost-pure- H_2 solid in a mixture of bulk composition 33 mol% O_2 at 296 K. At 7.9 GPa (a), the eutectic pressure, there is an equilibrium between solid H_2 (round darker region at top centre), the alloy (lower darker protrusion) and the fluid (light region). The ruby piece, used for pressure measurement, is touching the gasket on the mid-left giving a semicircular indentation in the crystal). At 8.4 GPa (b), the H_2 solid (light inverted pear-shaped region at top) and the alloy have grown to occupy the whole sample chamber.

Experiments performed at pressures of up to 10 bar have identified somewhat enigmatic temperature–pressure behaviour in these mixtures⁸ (Fig. 1). The solid line represents the explosion limit of nearly stoichiometric H_2/O_2 mixtures⁸. The so-called first, second and third explosion limits separate the regions of slow and fast reactions, and are associated with different mechanisms based on diffusion, heat release, mass-action kinetics and chain branching. Of particular interest here is the third explosion limit, which has been predicted^{8,9} to move to lower temperatures as pressure increases. This reaction should be thermodynamically favoured by the $\Delta V dP$ term, which makes the Gibbs free-energy change on combustion increasingly negative with increasing pressure (the volume of the product, H_2O (ref. 10) is always much smaller than the volume of the reactants H_2 (ref. 11) and O_2 (ref. 5), if they are ideally mixed). But we observed no crossing of an explosion limit on increasing the pressure: rather, almost complete stability of the mixtures is observed in the solid phase and the adjacent liquid phase. This leads us to suggest that the slope of the third explosion limit changes direction, becoming a fourth limit, and also that there exists a boundary at still higher pressures between slow and no reaction, which moves to lower pressures at lower temperatures. Indeed

we observe such a boundary in our low-temperature measurements down to 83 K. These two boundaries are indicated by dashed lines in Fig. 1 for a stoichiometric mixture. We propose that the upper boundary arises because at high densities a chain reaction cannot propagate as the active radical species (such as HO , HO_2 , H^* and O^*) cannot easily find acceptors to maintain or terminate the reaction when intermolecular distances decrease to values comparable to the dimensions of the molecules themselves. To go beyond this speculation would require much more experimental work.

We used microscopic and Raman observations to determine the binary phase diagram at 296 K. The boundary lines of phase transitions were determined by microscopic visual observation for various initial concentrations, as with our earlier investigations of the phase diagrams of H_2/He (ref. 12) and He/Ne (ref. 2) mixtures. Phase equilibria were clearly detected (Fig. 2). Raman measurements (performed using the 488-nm Ar^+ laser line at a power of 100 mW) gave complementary information. The molecular species present in the phases in equilibrium in the sample chamber could be identified from their vibron peaks. Also, analysis of the intensity of the H_2 and O_2 vibron peaks and their ratio could probe the homogeneity and the approximate

FIG. 3 Pressure dependence (at 296 K) of the vibron frequency of H_2 (a), and of O_2 (b), in O_2/H_2 mixtures, with O_2 molar concentrations X of 0, 25, 33, 50, 75 and 100 mol%. The vertical dashed line indicates the pressure of the eutectic point above which a solid–solid separation of phases is observed, giving almost-pure H_2 , almost-pure O_2 and an alloy. The solid lines represent our measurements of the vibron frequency in pure H_2 and in pure O_2 . The open circles are our data points in the O_2/H_2 alloy.



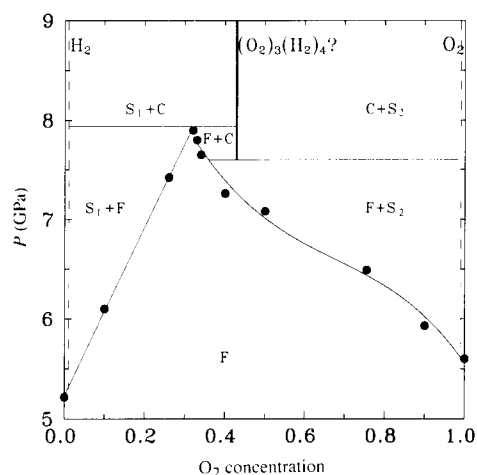


FIG. 4 O_2/H_2 binary phase diagram at 296 K. The filled circles correspond to our measurements of the solid + fluid $\rightarrow$ fluid ($\text{S} + \text{F} \rightarrow \text{F}$) transition, and the interpolated solid line indicates the $\text{S} + \text{F}$ boundary. The vertical dashed lines are the stability limits of a hydrogen-rich solid, S_1 , and an oxygen-rich solid, S_2 ; the vertical thick line indicates the stability domain of the alloy, possibly the compound $(\text{O}_2)_3(\text{H}_2)_4$, C.

proportion of the solid phases in equilibrium. More quantitatively, in the fluid phase, the dependence of the H_2 vibron frequency on O_2 concentration (Fig. 3a) and the dependence of the O_2 vibron frequency on H_2 concentration (Fig. 3b) were measured for calibration purposes (to check the constancy of concentrations in the sample chamber with time after successive pressure cycles and between different loadings).

Total miscibility of O_2 and H_2 in the fluid phase is observed. As the pressure is increased, the fluid separates in a solid–fluid equilibrium. The solid + fluid $\rightarrow$ fluid transition line was measured with good reproducibility at various concentrations (Fig. 4). Its shape is characteristic of a eutectic binary phase diagram, with an eutectic point around 32 mol% O_2 and at 7.9 GPa. On increasing the pressure further, various solid–solid phase separation processes were observed, depending on the initial concentration. A mixture with 10 mol% O_2 separates above 8 GPa into almost pure solid H_2 (in which no O_2 vibron peak could be detected) and a small region of a distinct phase; for a mixture with 90 mol% O_2 , an equilibrium exists between almost pure solid O_2 (no H_2 vibron peak) and another phase. The evolution of the nearly eutectic concentration is illustrated in Fig. 2. At 7.9 GPa (the eutectic pressure) an equilibrium between two solids and a fluid is observed (Fig. 2a). As pressure increases, the two solids grow homogeneously to occupy the whole sample chamber (Fig. 2b). In the smaller solid region only the vibron peak of H_2 could be measured, but in the other solid region the vibron peaks of both O_2 and H_2 were quite intense, with an intensity ratio independent of the position of the laser spot through it. Hence, the small solid region is almost pure hydrogen, whereas the other is an O_2/H_2 alloy. Its proportion can be estimated by the lever rule from the amounts of the two phases in equilibrium, under the assumption of ideal mixing using the equations of state of solid hydrogen¹¹ and solid oxygen^{5,13}. From Fig. 2b, we estimate that the alloy phase contains 43 mol% O_2 ; averaging over other samples, the composition of the alloy was estimated to be 42 ± 1 mol% O_2 . This is confirmed by the observations of solid–solid phase separation in initial mixtures slightly differing from 42 mol% O_2 . For initial O_2 molar concentrations of 40, 42 and 50 mol% and pressures above 8 GPa, the alloy coexists respectively with a small amount of solid H_2 , exists in almost pure form, and coexists with solid O_2 . Furthermore, for O_2 concentrations greater than 35 mol% in the initial mixture, the almost-pure O_2 solid + H_2 -rich fluid equilibrium (first observed by going up in pressure), transforms above 7.6 GPa to a new phase separation, either between the O_2/H_2 alloy and H_2 -rich fluid phase or between the alloy and almost pure solid O_2 , depending whether or not the initial concentration is greater than 42 mol% O_2 .

These observations lead us to propose the binary phase diagram presented in Fig. 4, which indicates that a O_2/H_2 alloy of

O_2 concentration around 42 mol% is stable above 7.6 GPa. At this stage, we cannot tell whether this alloy is a disordered solid solution or a stoichiometric compound; the latter would have the composition $(\text{O}_2)_3(\text{H}_2)_4$. Single-crystal synchrotron X-ray diffraction (the required technique for characterizing the structure of low-atomic-weight solid in a diamond-anvil cell) would provide structural information on this alloy. We tried such an experiment twice at the ESRF synchrotron, but unfortunately after some time in the beam a combustion reaction was induced. The Raman vibron shifts for H_2 and O_2 are almost identical to those in the bulk solids, which is quite surprising in view of the large concentration effect observed in the fluid phase; this seems to be a coincidence, which may reflect certain restrictions on the local environment of the molecules.

We probed the evolution of the binary phase diagram at low temperature by measuring phase transitions in a mixture containing 33 mol% O_2 . We observed a eutectic P – T line and the stability of the alloy at lower density (at 83 K above pressures of 0.5 GPa). Concentrations of the phases in equilibrium were estimated either from the Raman spectra as described above or from the ratio of the volumes of the equilibrium phases. Two trends are seen which are consistent with common rules for binary systems¹⁴. First, the eutectic trough deepens with increasing pressure. Second, the eutectic moves toward the component with the lowest melting temperature: the O_2 concentration of the eutectic decreases by a few per cent between 296 K and 83 K.

This new physical chemistry of dense O_2/H_2 might prove useful for applications such as rocket-fuel or energy-storage technologies. It could also be relevant to models of the interiors of the jovian planets and their satellites¹⁵, which currently assume that total combustion of oxygen has taken place and which are based on the phase diagrams of mixtures of just hydrogen, helium and ices (H_2O , NH_3 , CH_4). □

Received 9 January; accepted 25 September 1995.

- Vos, W. L. *Nature* **358**, 46–48 (1992).
- Loubeyre, P., Jean-Louis, M., LeToullec, R. & Charon-Gérard, L. *Phys. Rev. Lett.* **70**, 178–181 (1993).
- Loubeyre, P., LeToullec, R. & Pinceaux, J. P. *Phys. Rev. Lett.* **72**, 1360–1363 (1994).
- Mao, H. K. & Hemley, R. *Rev. mod. Phys.* **66**, 671–692 (1994).
- Rooff, A. L. & Desgreniers, S. in *Molecular Systems Under High Pressure* (eds Pucci, R. & Picot, G.) 123–138 (North-Holland, Amsterdam, 1991).
- LeToullec, R., Pinceaux, J. P. & Loubeyre, P. *High Press. Res.* **1**, 77–90 (1988).
- Mao, H. K., Xu, J. & Bell, P. M. *J. Geophys. Res.* **91** (B5), 4673–4676 (1986).
- Dougherty, E. P. & Rabitz, H. *J. chem. Phys.* **72**, 6571–6586 (1980).
- Astapov, N. S. & Nikolaev, Y. A. *Fizika Goreniya Vzryva* **20**, 89–96 (1984).
- Hemley, R. J. *et al. Nature* **330**, 737–740 (1987).
- Hemley, R. J. *et al. Phys. Rev.* **B42**, 6458–6470 (1990).
- Loubeyre, P., LeToullec, R. & Pinceaux, J. P. *J. condens. Matter* **3**, 3183–3192 (1991).
- Olinger, B., Mills, R. L. & Roof, R. B. *J. chem. Phys.* **81**, 5068–5073 (1984).
- Young, D. A. *J. chem. Phys.* **98**, 9819–9829 (1993).
- Hubbard, W. B. *Planetary Interiors* (Van Nostrand, New York, 1984).

ACKNOWLEDGEMENTS. We thank D. Hausermann and M. Hanfland for help with synchrotron measurements at ESRF.

Lamellar aluminophosphates with surface patterns that mimic diatom and radiolarian microskeletons

Scott Oliver*, Alex Kuperman*, Neil Coombs†, Alan Lough* & Geoffrey A. Ozin*‡

* Materials Chemistry Research Group, Lash Miller Chemical Laboratories, University of Toronto, 80 St George Street, Toronto, Ontario M5S 1A1, Canada

† Imagetek Analytical Imaging, 32 Manning Avenue, Toronto, Ontario M6J 2K4, Canada

ORGANISMS such as diatoms and radiolaria synthesize elaborate biomineral exoskeletons which display hierarchical structures patterned on length scales from less than a micrometre to millimetres. Synthetic materials chemistry, in contrast, has traditionally been able to achieve regular patterning only on microscopic ($<10\text{ Å}$) and more recently^{1–3} mesoscopic ($10\text{--}10^3\text{ Å}$) length scales. Here we report the synthesis of crystalline, lamellar aluminophosphate structures that are patterned on the submicrometre-to-millimetre scales found in the living world. As in the syntheses of ordered mesoporous solids^{1–3}, our approach involves templating by self-assembled organic aggregates, and we propose that the larger scale of the patterning here arises from the involvement of vesicle templates rather than the micelle-like or bilayer structures thought to be responsible for mesoscale pattern formation^{1–3}.

Our synthesis system makes use of a non-aqueous tetraethylene glycol (TEG) solvent, into which phosphoric acid (85 wt%) and an amphiphilic alkylamine, here $\text{C}_{10}\text{H}_{21}\text{NH}_2$, are added. Gelation occurs immediately, yielding a viscous, opaque, white suspension of a finely crystalline mesolamellar alkylammonium dihydrogen phosphate phase (see below). Pseudo-boehmite (Dispal 23N4-80, Vista) is then added to the synthesis mixture, giving a gel composition of 14 TEG:0.9 $\text{Al}_2\text{O}_3 \cdot 2.5\text{H}_2\text{O}$:1.8 P_2O_5 :3.0 $\text{C}_{10}\text{H}_{21}\text{NH}_2$. The gel is loaded into Teflon-lined stainless steel autoclaves and treated at 180°C for 72 hours. The products are filtered, washed with copious amounts of deionized water and acetone, and allowed to dry in ambient air.

We recrystallized the decylamine-phosphoric acid mixture in water at 50°C , producing large crystals of decylammonium dihydrogen phosphate, $(\text{C}_{10}\text{H}_{21}\text{NH}_3^+)(\text{H}_2\text{PO}_4^-)$, denoted DDP. The crystal structure was solved by single-crystal X-ray diffraction (space group $P2_1/n$, unit-cell dimensions $a=7.358(2)\text{ Å}$, $b=39.749(12)\text{ Å}$, $c=9.131(4)\text{ Å}$, $\beta=90.45(1)^\circ$; see Supplementary Information) and consists of hydrogen-bonded layers of dihydrogen phosphate anions, between which the charge-balancing surfactant cations reside. In this structure, the hydrophobic tails are interdigitated and orthogonal rather than lying in the more common bilayer arrangement. An extensive hydrogen-bonding network exists between the protonated surfactants and dihydrogen phosphate counter-anions.

The powder X-ray diffraction (PXRD) pattern of the starting synthesis mixture before thermal treatment is shown in Fig. 1a. This PXRD pattern can be indexed to the DDP lamellar mesophase. Also present in the PXRD pattern are the low-intensity, broad diffraction peaks of the pseudo-boehmite starting material. The PXRD pattern of the final product, Fig. 1b, reveals the presence of three phases. There exist only trace amounts of the starting DDP phase and AlPO_4 -cristobalite. The predominant phase is a mesolamellar aluminophosphate, denoted MLA, with an interlamellar spacing of $\sim 29.28\text{ Å}$. The first three intense ($00l$) peaks ($l=1, 2, 3$) are evident and the phase gives no higher-angle peaks. This is diagnostic of an inorganic-organic compos-

ite mesolamellar phase, lacking long-range order within the layer regions of the structure.

The 29.28-Å aluminophosphate mesophase is comprised of two dominant morphologies. One of these is a featureless, flakey morphology. The other has the form of spheres $1\text{--}2\text{ mm}$ in diameter (Fig. 2a). These spheres represent $\sim 10\text{ wt\%}$ of the product and consist exclusively of the 29.28-Å MLA phase—they contain no DDP or dense-phase AlPO_4 -cristobalite as evidenced by PXRD and thermogravimetric analysis coupled mass spectroscopy (TGA-MS) data. The spheres are easy to handle but excessive pressure breaks them into plate-like remnants.

The MLA materials are stable in the beam of an electron microscope. Quantitative energy dispersive X-ray fluorescence (EDX), as well as elemental inductively coupled plasma (ICP) analyses show them to have Al:P ratios of $\sim 1:2$. The organic content is typically $40\text{--}50\text{ wt\%}$, corresponding to a phosphorus to alkylamine mole ratio of $\sim 1:1$. We have obtained a transmission electron micrograph of the mesolamellar structure (not shown), which confirms the approximate lamellar spacing of 30 Å .

TGA-MS data of the spheres and flakes display two thermal events, at around 200 and 350°C . They correspond to loss of surfactant, approximating 35 and $15\text{ wt\%}$ of the solid, respectively. The PXRD pattern of the sample treated at 200°C shows that the 29.28-Å mesolamellar aluminophosphate phase has transformed to one with a lamellar spacing of 23.50 Å , presumably corresponding to loss and reorganization of interlayer surfactant. Scanning electron microscopy (SEM) studies show severe deformation of the surface patterns. Harsher treatment at 300°C completely removes the organics, causing the mesolamellar structure to collapse to AlPO_4 -cristobalite with complete loss of the surface patterns.

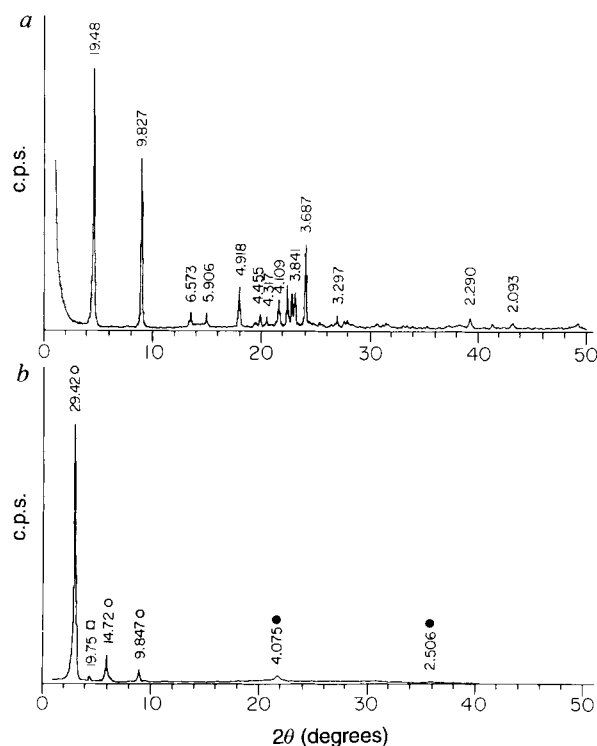


FIG. 1 a, Powder X-ray diffraction (PXRD) pattern of the synthesis gel before thermal treatment. The pattern indexes well to the decylammonium dihydrogen phosphate SC-XRD, single-crystal X-ray diffraction, structure. b, PXRD of the synthesis product after treatment at 180°C for 72 h. Open circles, the mesolamellar aluminophosphate (MLA) majority phase; open squares, trace amounts of decylammonium dihydrogen phosphate; filled circles trace amounts of AlPO_4 -cristobalite. (c.p.s., counts per second.)

* To whom correspondence should be addressed.

SEM images (Fig. 2) of the synthesis products reveal surface structural features that span the entire submicrometre to millimetre range. Figures 2*b–d* show close-up views of various areas of the sphere surface. The higher-magnification micrograph, Fig. 2*b*, shows an area containing an extraordinary surface morphology of close-packed bowl-like impressions, varying in diameter from several micrometres to $>100\text{ }\mu\text{m}$. Other areas of the sphere surface, Fig. 2*c*, contain arrays of similarly packed bowl-shaped features, but in this case with a concentric arrangement of platelets growing out of the interior of the bowl, together with a finer structure of narrow ridges and submicrometre pores. Still other areas possess a remarkable honeycomb-like pattern of $1\text{-}\mu\text{m}$ -sized bowls of roughly hexagonal symmetry, Fig. 2*d*.

We propose that the observed MLA morphologies and surface patterns are templated by two types of DDP assemblies, in equilibrium under the given synthesis conditions. Non-aqueous glycol-based solvents containing surfactants are known to be microstructured^{4,5}. The crystalline DDP lamellar phase initially present should also exist in the TEG solvent, as the synthesis gel is increased in temperature. The packing requirements of the

small ammonium surfactant head group favours a planar bilayer structure⁶. Such a planar bilayer template would give rise to the flakey morphology MLA phase, with no surface patterning. The TEG, however, can be expected to function as both solvent and as a co-surfactant. Incorporation of TEG molecules into the DDP bilayers would introduce curvature into the assembly⁷ leading to vesicle formation. Such bilayer-to-vesicle transformations are well documented^{8,9}. In addition, mineralized aluminium(III) species can trigger phase separation (that is, demixing) of the $\text{C}_n\text{H}_{2n+1}\text{NH}_3^+$ ionic surfactant and TEG non-ionic co-surfactant¹⁰. We propose that ordering of these domains might lead to patterns on the surface of the vesicles. Deposition of aluminophosphate lamellae will occur preferentially in the $\text{C}_n\text{H}_{2n+1}\text{NH}_3^+\text{H}_2\text{PO}_4^-$ domains of the vesicle. Both before and after washing the material with water, we observe 'replicated' patterns of submicrometre surface pores, Fig. 2*c*.

We suppose that adhesion of vesicles on the surface of a growing aluminophosphate sphere directs the deposition of the bowl-shaped features with superimposed fine patterning, Fig. 2*b*. The packing of these vesicles will control the overall surface bowl

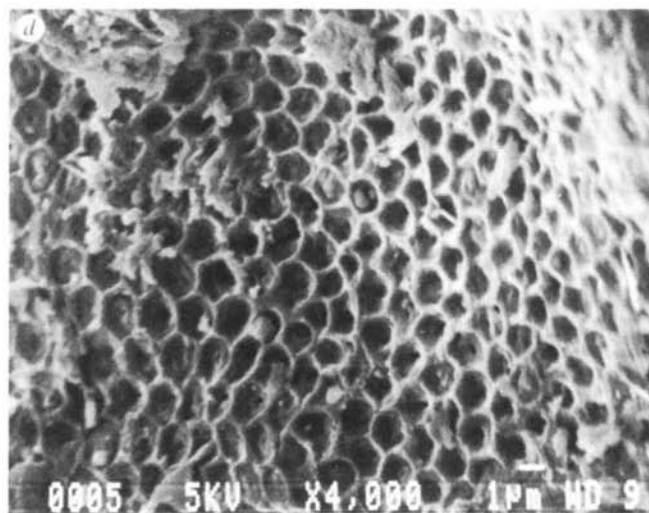
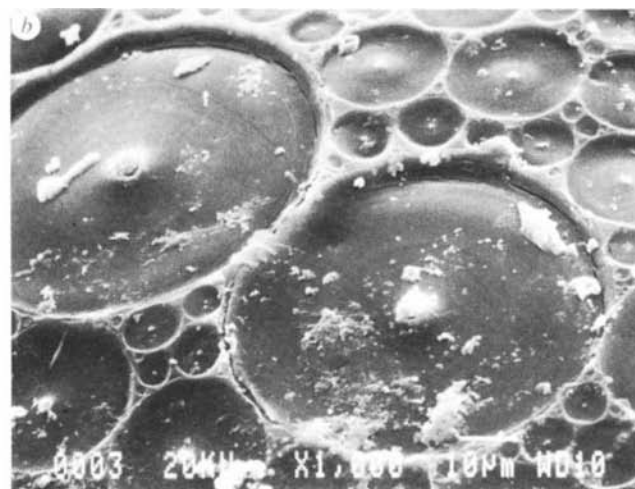
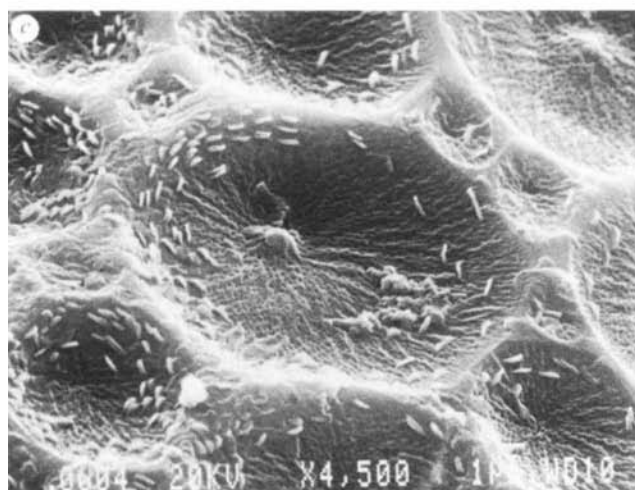


FIG. 2 Scanning electron micrographs of the decylamine mesolamellar aluminophosphate (MLA) product. *a*, Overview of an MLA sphere. *b*, Higher magnification of the sphere, showing the $1\text{--}50\text{ }\mu\text{m}$ diameter surface bowls. Note the packing of the surface bowls. *c*, The interiors of some of these surface bowls contain a ridged fine-structure and submicrometre pores, as well as a concentric arrangement of platelets. *d*, $1\text{-}\mu\text{m}$ honeycomb surface morphology. Scale bars: *a*, 0.60 mm ; *b*, $10\text{ }\mu\text{m}$; *c*, *d*, $1\text{ }\mu\text{m}$.

FIG. 3 Schematic illustration of proposed patterning mechanism. The circles with tails represent the surfactant cations, the crosses represent the dihydrogen phosphate counter-anions, and the connected circles represent the TEG molecules. *a*, Planar interdigitated lamellar solid phase; *b*, alkylammonium dihydrogen phosphate bilayer existing in the TEG solvent; *c*, alkylammonium dihydrogen phosphate vesicle coexisting in the TEG solvent. The vesicle contains surface TEG molecules that introduce curvature; *d*, aluminophosphate polymerization of the bilayer and formation of the flakey mesolamellar aluminophosphate (MLA) morphology; *e*, aluminophosphate polymerization of the vesicle and formation of the spherical MLA morphology, with its surface patterning. (SC-XRD, single-crystal X-ray diffraction.)

Mesolamellar aluminophosphates

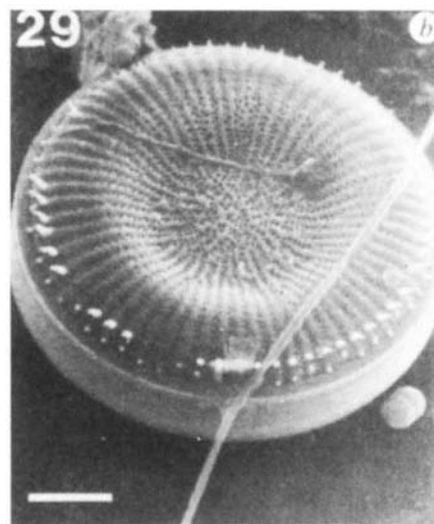
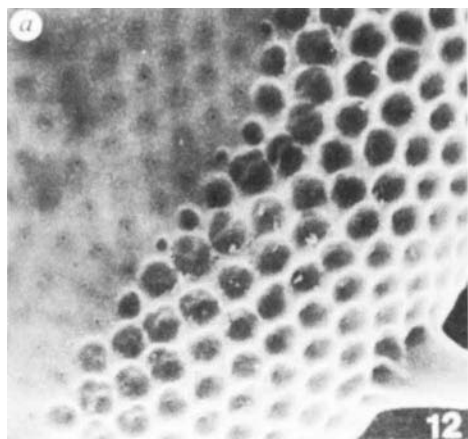
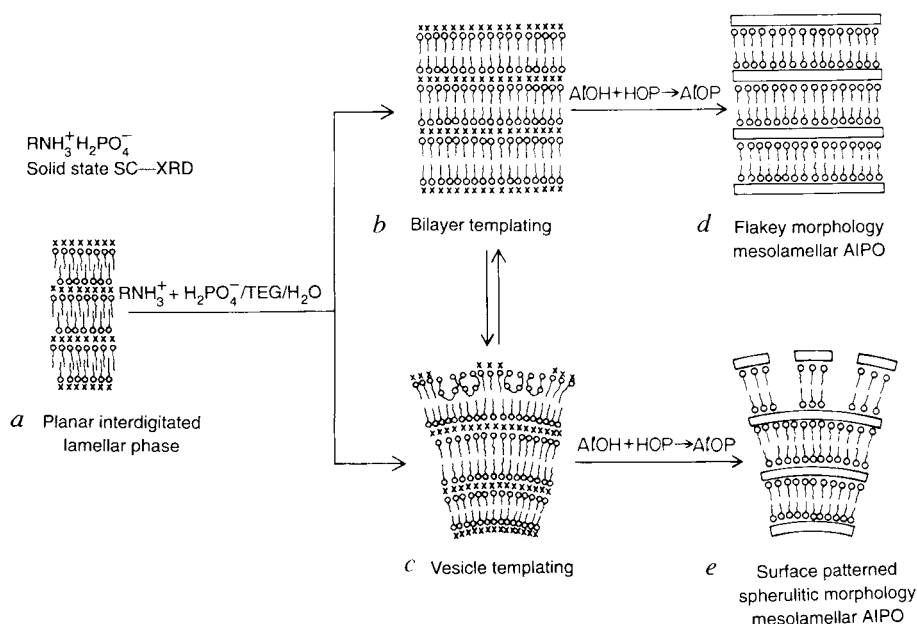


FIG. 4 SEM images of examples of the skeletal architectures of radiolaria and diatoms. *a*, *Challengerosium avicularioides* ($\times 500$). Note the striking similarity to the synthetic structure of Fig. 2*d*. *b*, *Stephanodiscus* (scale bar, 10 μm). Discoid-shaped structure displaying rings of crystal-like protrusions of a genre similar to the synthetic structure of Fig. 2*c*. *c*, *Protocystis murrayi* ($\times 1,800$). Circular surface bowl-like depressions are common among the protozoan skeletons. Note the similarity to the synthetic structure of Fig. 2*b*. Reproduced from references 14, 15 with permission.

architecture. Minimization of their surface free energy through a balance of adhesion and elastic forces⁶ could lead to vesicle distortion and to the honeycomb surface patterns. Fig. 2c, d. This tentative model is represented in Fig. 3.

Specific control experiments have cast further light on the origin of the observed morphologies and patterns. TEG allows formation of the surface patterns, whereas ethylene glycol and higher-molecular-weight polyethylene glycols do not. Increasing the mole ratio of water, varying the mole ratio of surfactant or changing the temperature, allow formation of a mesolamellar aluminophosphate, but devoid of surface patterning. A small-molecule amine unable to form self-assembled aggregates results in normal faceted crystal habits of one-dimensional (chain), two-dimensional (porous sheet) and three-dimensional (open-framework) aluminophosphate materials¹¹. These experiments emphasize the synergistic role of the surfactant-co-surfactant templates in determining morphology and surface patterning in the MLA synthesis system.

Our model for pattern formation is akin to some of those proposed in the biomineralization literature for the creation of exoskeletons^{12,13}. One prevailing idea is that radiolaria skeletons grow by the secretion of silica into a network of bubble-like alveoli^{12–16}. This model has features in common with our growing aluminophosphate sphere covered by vesicles. The concentric nature of the vesicles could facilitate the deposition of replica features on the surface of the bowls. This could account for the concentric ring and platelet formations in the bowls of Fig. 2c.

It is interesting to compare the morphologies and patterns of the artificial inorganic assemblies (Fig. 2c, d) with those of the siliceous skeletons of diatoms and radiolaria (Fig. 4a, b).^{14–16} The skeletal features are of similar size and form to those of the surfactant-based aluminophosphate morphologies. Examples of surface bowl-like features akin to the synthetic ones in Fig. 2b are common in radiolaria, Fig. 4c. Another common feature is the lack of long-range order in the aluminophosphate regions

of the synthetic ultrastructures, a situation similar to the amorphous silica protozoan exoskeletons of radiolaria and diatoms¹⁷, as well as the amorphous walls of micelle-templated mesoporous silicas^{1–3}.

It would seem that there are interesting parallels to be drawn between protozoan diatoms and radiolaria and our artificial inorganic assemblies. The inorganic materials reported here are hierarchical and display structural features on length scales ranging from submicrometre to millimetres. The ability to synthetically control macroscopic form and surface patterns of inorganic materials could find important application in areas such as catalysis, separation technology and the development of new functional materials¹⁸. □

Received 15 May; accepted 12 October 1995.

1. Kresge, C. T. et al. *Nature* **359**, 710–712 (1992).
2. Beck, J. S. et al. *J. Am. chem. Soc.* **114**, 10834–10843 (1992).
3. Monnier, A. et al. *Science* **261**, 1299–1303 (1993).
4. Kaler, E. W. & Martino, A. *Langmuir* **11**, 779–784 (1995).
5. Thomas, B. N., Safinya, C. R., Plano, R. J. & Clark, N. A. *Science* **267**, 1635–1638 (1995).
6. Israelachvili, J. N. *Intermolecular and Surface Forces* 2nd edn (Academic, New York, 1991).
7. Hoffmann, H. *Adv. Mater.* **6**, 116–129 (1994).
8. Haslop, W. P., Allonby, J. M., Akred, B. J. & Messenger, E. T. US Patent No. 4,618,446 (1986).
9. Akred, B. J., Messenger, E. T. & Nicholson, W. J. US Patent no. 4,659,497 (1987).
10. Fendler, J. H., *Membrane Mimetic Chemistry* (Wiley, New York, 1982).
11. Oliver, S. et al. *Stud. Surf. Sci. Catal.* **84**, 219–225 (1994).
12. Simkiss, K. & Wilbur, K. M. *Biomineralization: Cell Biology and Mineral Deposition* (Academic, San Diego, 1989).
13. Mann, S., Webb, J. & Williams, R. (eds) *Biomineralization: Chemical and Biochemical Perspectives* (VCH, New York, 1989).
14. Takahashi, K. (ed) *Radiolaria* (Ocean Biocoenosis Ser. no. 3, Woods Hole Oceanographic Institution, Woods Hole, MA, 1991).
15. Richard, M. (ed.) *Ouvrage Dedicé à la Mémoire du Professeur Henry Germain* (Koeltz Scientific, Königstein, 1990).
16. Anderson, O. R. *Radiolaria* (Springer, New York, 1983).
17. Mann, S. *J. mater. Chem.* **5**, 935–946 (1995).
18. Oliver S. & Ozin, G. A. *Adv. Mater.* (in the press).

SUPPLEMENTARY INFORMATION. Crystallographic refinement data are available on request. Contact Mary Sheehan at the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. This work was supported by the Natural Sciences and Engineering Research Council of Canada. S.O. was supported in part by the Ontario Graduate Scholarship program.

High concentrations and photochemical fate of oxygenated hydrocarbons in the global troposphere

Hanwant B. Singh*, M. Kanakidou†, P. J. Crutzen‡ & D. J. Jacob§

* NASA Ames Research Center, Moffett Field, California 94035, USA

† Centre des Faibles Radioactivités, CNRS-CEA, F-91198 Gif-sur-Yvette Cedex, France

‡ Max Planck Institute for Chemistry, D-55020 Mainz, Germany

§ Harvard University, Cambridge, Massachusetts 02138, USA

OXYGENATED species in the atmosphere are important sources of free radicals and are intricately linked with the fate of nitrogen oxides (NO_x), which are themselves necessary for tropospheric ozone formation^{1,2}. With the exception of formaldehyde, oxygenated hydrocarbons have rarely been measured in the free troposphere. Here we report airborne measurements indicating the presence of high concentrations (compared to those of routinely measured C_2 – C_6 tropospheric hydrocarbons^{3,4}) of acetone and methanol. We use a three-dimensional model to show that acetone photochemistry provides a quantitatively significant (up to 50%) pathway for sequestering NO_x in the form of peroxyacetyl nitrate, a relatively unreactive temporary reservoir of NO_x . Furthermore, in the dry regions of the upper troposphere, acetone can provide a large primary source of HO_x ($\text{OH} + \text{HO}_2$) radicals, resulting in

increased ozone production. This surprisingly significant contribution of such oxygenated hydrocarbons to tropospheric NO_x , HO_x and ozone cycling is likely to be affected by their changing natural and anthropogenic emissions due to land-use change, biomass burning and alcohol-based biofuel use.

Limited measurements of alcohols and carbonyls have been reported from ground sites generally in rural/urban environments^{5–7}. Recently, we have developed and tested a semi-automated airborne instrument that uses a Reduction Gas Detector (RGD) for the sensitive (~ 10 parts per trillion (10^{12}) by volume, p.p.t.) detection of carbonyls in real time^{8,9}. We have now expanded this to include alcohols. Figure 1 shows a chromatogram that demonstrates the separation and detection of acetaldehyde, acetone, methanol and ethanol from free tropospheric air. A parallel instrument also measured peroxyacetyl nitrate (PAN) to a sensitivity of ~ 1 p.p.t. Data were collected in February–March 1994 during the NASA Pacific Exploratory Mission, PEM-West (B); an airborne campaign to study the atmosphere over the Pacific Ocean.

Figure 2 shows the latitudinal distribution of acetone and PAN in the free troposphere (5–10 km height) based on a series of six flights over the Pacific Ocean starting from 40°N to 10°S over a longitude of 125°W to 140°E . During these selected missions, westerly air flows and the great distance from the Asian continent precluded a direct effect of continental emissions, and these air samples should represent near-background conditions for this season. In the free troposphere, acetone concentrations of the order of 500 p.p.t. were present at northern mid-latitudes and declined to ~ 200 p.p.t. at southern latitudes. These mixing ratios are similar to those measured in the Canadian Arctic/subarctic when clean conditions prevailed⁹, but are higher than

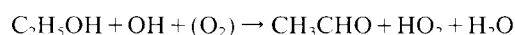
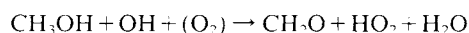
those reported by Arnold *et al.*¹⁰ for the mid-latitude troposphere (44° N; February May) over Europe using an indirect method involving chemical ionization mass spectrometry. PAN concentrations in the vicinity of 100 p.p.t. were present in the northern free troposphere, although its abundance was highly variable. Near the tropics, mixing ratios below 10 p.p.t. were often seen. PAN exhibited a strong vertical gradient, and its concentrations in the marine boundary layer were typically <2 p.p.t. Acetaldehyde could not be rigorously quantified because of difficulties involving its calibration. Based on relative response factors determined in the laboratory, it would seem that 30–100 p.p.t. of acetaldehyde were nearly always present.

In the free troposphere ~700 p.p.t. of methanol were present at northern mid-latitudes, declining to ~400 p.p.t. at southern latitudes (Fig. 3). In general, ethanol abundance was an order of magnitude lower than that of methanol. There were instances when ethanol was below the detection limit of 10 p.p.t. No latitudinal profiles for alcohols or acetone have been reported in the

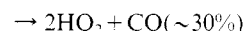
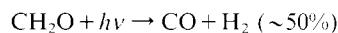
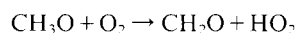
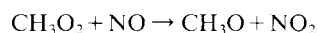
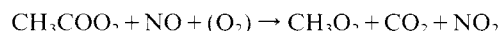
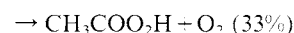
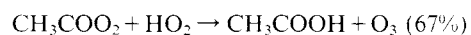
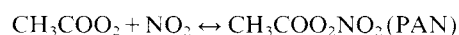
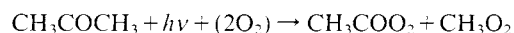
literature for comparison. Limited measurements were also made in the mid-latitude lower stratosphere (331±104 p.p.b. (10⁻⁹ by volume) O₃; -55 to -69 °C dew point). Under these conditions, mixing ratios of acetone and methanol were 148±63 p.p.t. and 95±51 p.p.t. respectively, while ethanol was generally below the detection level.

The main removal process for acetone, methanol and ethanol is expected to be reaction with OH radicals and photolysis (important for acetone only). Based on published OH radical rate coefficients¹¹, the best estimates for OH radical abundance (10⁶ molecules per cm³; methyl chloroform lifetime of 5.0 yr), and known photolytic cross-sections, we estimate that average lifetimes of acetone, methanol and ethanol are of the order of 16 days, 16 days and 4 days, respectively. To balance the measured abundances and the removal rates, a global source of about 50 Tg yr⁻¹ acetone, 45 Tg yr⁻¹ methanol and 15 Tg yr⁻¹ ethanol is required. Although these species are miscible with water, their Henry's Law coefficients are relatively low (at 25 °C; 32 M atm⁻¹ for acetone; 220 M atm⁻¹ for alcohols) and removal by wash-out/rainout processes or the oceanic sink should not compete favourably with removal by reaction with OH radicals and photolysis¹². Upper-tropospheric cloud systems may take in some alcohols, but are likely to desorb them during precipitation. The north-south latitudinal gradient implies larger sources in the north, but we note that photochemical removal at this time of the year (February) is about 1.5–2 times higher in the Southern Hemisphere than the Northern Hemisphere. Additional gradients between the marine and continental environments may also exist.

Alcohols on further oxidation form photochemically active carbonyls and free radicals¹³.



Acetone also can easily react with OH radical and be photolysed to produce free radicals (for example, HO₂) and several stable species such as PAN, and acetic and peracetic acids^{9,14}.



We have studied the role of acetone in PAN formation by incorporating detailed acetone and C₁–C₃ hydrocarbon chemistry in a three-dimensional model of the atmosphere¹⁴. Based on the inventory suggested by Singh *et al.*⁹, a 49 Tg yr⁻¹ global source of acetone was incorporated in this model. Emissions from vegetation were scaled to the annual mean net primary productivity, while anthropogenic emissions were distributed in a manner similar to those of CO₂. This resulted in acetone concentrations shown in Fig. 4a. Model PAN concentrations were calculated with and without the inclusion of acetone. Figure 4b shows the PAN mixing ratios that could be attributed exclusively to acetone chemistry. It is evident that acetone is responsible for sizeable (10–45 p.p.t.) concentrations of PAN in the free troposphere. Its effect is greatest in the middle and upper troposphere where photolysis of acetone is quite efficient and PAN is

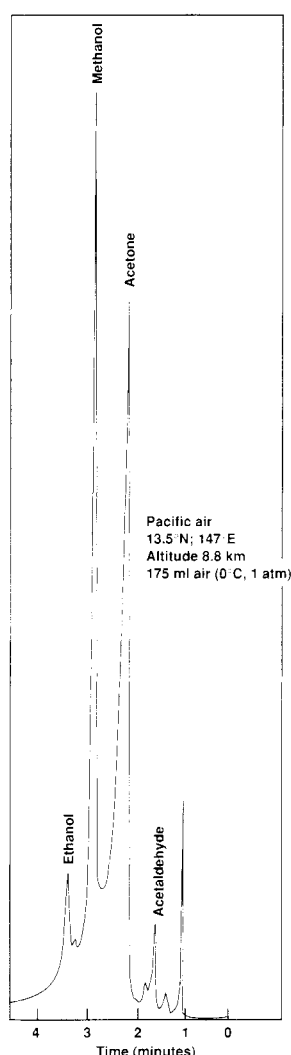
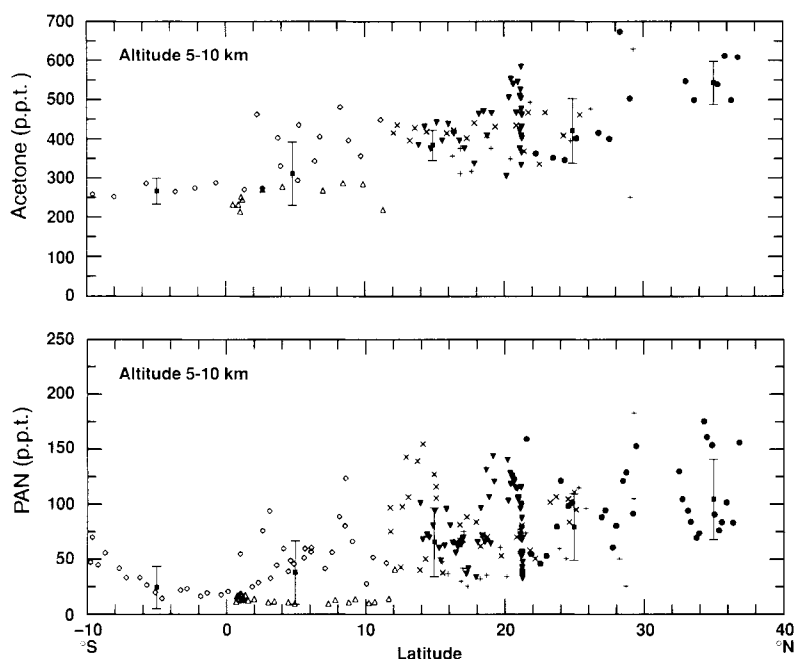


FIG. 1 Chromatogram showing the detection and separation of selected oxygenated species from free tropospheric air. A 175-ml sample of air is freeze trapped at -140 °C in a sample loop, and its constituents separated on a 30-m stripper (for water removal) and a 60-m analytical quartz capillary column (0.53 mm i.d., 1 µm DB wax film coating; 80 °C) and detected with a Reduction Gas Detector (RGD). A heated (20 °C) Teflon probe is used for sampling, and all calibrations, using permeation tubes held at 0 °C, are done in-flight in a manner that mimics ambient air sampling. The tailing of the acetone peak is probably due to its slow release from the mercuric oxide bed in the RGD.

FIG. 2 Latitudinal distribution of acetone (top panel) and PAN (bottom panel) in the free troposphere over the Pacific. These data are for 5–10 km altitude, and are based on a series of six aircraft flights performed during 7–19 February, 1994. Each flight is represented by a separate symbol. Transit flights started from California (38° N) to Hawaii (21° N) (filled circles) and Hawaii to Guam (13° N) (inverted filled triangles). Four additional flights were undertaken north (+ and x) and south (○, △) from Guam. Each filled square and the line extending from it shows mean ± 1 standard deviation for data within the 10° latitude belt.



most stable. Based on model results and comparison with data, up to 50% of observed PAN may be formed by this mechanism. Model results also show that acetone photolysis can produce 10^5 – 10^6 molecules per cm^3 of peroxyacetyl radicals with highest values in the tropics. We suggest that acetone provides an important pathway in storing NO_x in the form of PAN, and must be included in global models that aim to accurately simulate ozone in the troposphere.

We also find that acetone is an important and previously unrecognized source of HO_x radicals in the upper troposphere. Current photochemical models identify the two main upper-tropospheric sources of HO_x as the reaction of $\text{O}(^1\text{D})$ with water vapour ($\text{O}_3 + h\nu \rightarrow \text{O}(^1\text{D}) + \text{O}_2$; $\text{O}(^1\text{D}) + \text{H}_2\text{O} \rightarrow 2\text{OH}$) and the photolysis of CH_2O . The main source of CH_2O in these models is the oxidation of methane by OH radicals; therefore one must view the $\text{O}(^1\text{D}) + \text{H}_2\text{O}$ reaction as the primary source of HO_x

and CH_2O as an amplifier of this source. Under dry conditions of the upper troposphere, where the reaction $\text{O}(^1\text{D}) + \text{H}_2\text{O}$ is relatively slow, acetone makes an important additional contribution to the primary source of HO_x . The photolysis of acetone here yields two HO_2 and two CH_2O molecules ($\text{NO} \gg \text{HO}_2$), and 30% of the CH_2O molecules photolyse via the radical branch to yield two more HO_2 molecules. Thus the HO_x yield from the photolysis of acetone is ~ 3.2 ($2 + 4 \times 0.30$), as compared to a yield of 2 from the reaction of $\text{O}(^1\text{D}) + \text{H}_2\text{O}$. In a simple photochemical model calculation for the upper troposphere at the equinox (40° N, 11 km, 50 p.p.b. O_3 , 90 p.p.m. (10^{-6} by volume) H_2O and 0.5 p.p.b. acetone), we find 24-hour average HO_x production rates of 7×10^3 molecules per cm^3 per s from reaction of $\text{O}(^1\text{D}) + \text{H}_2\text{O}$ and 9×10^3 molecules per cm^3 per s from photolysis of acetone. When all primary and secondary sources are included, acetone may enhance total HO_x concentrations by up

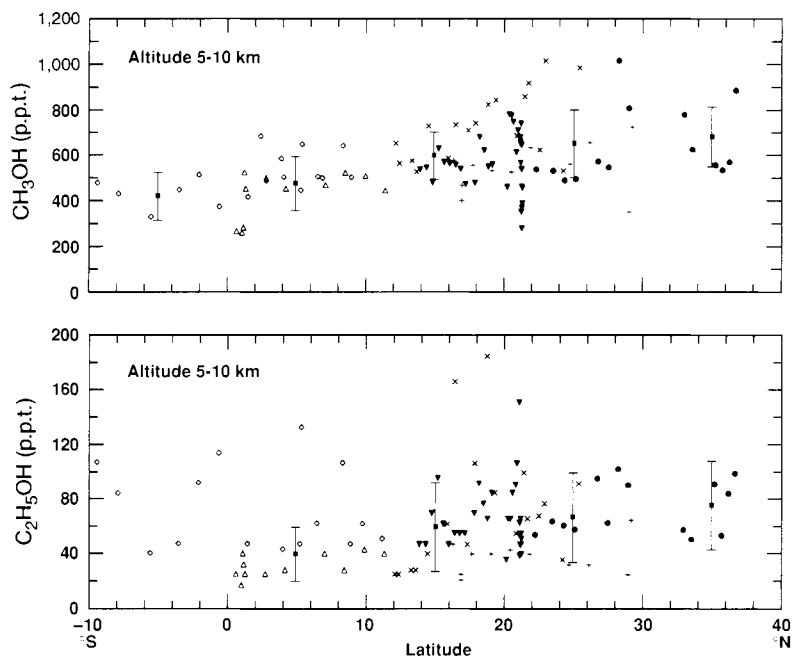
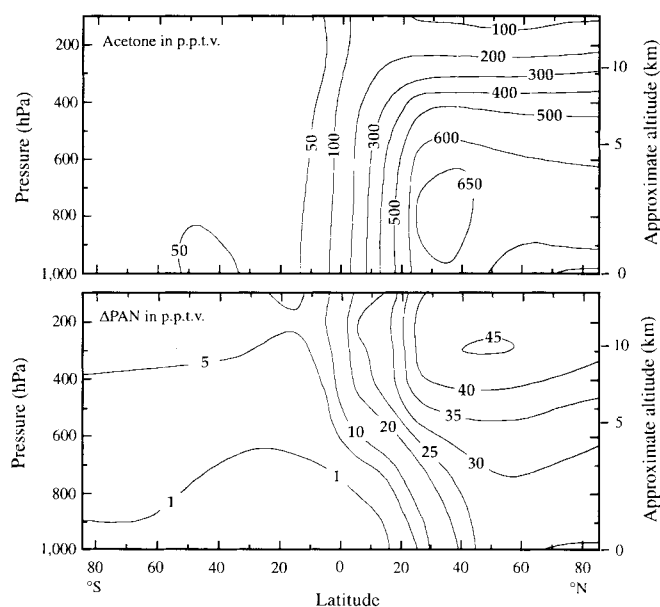


FIG. 3 Latitudinal distribution of methanol (top panel) and ethanol (bottom panel) in the free troposphere over the Pacific Ocean. Symbols are as in Fig. 2.

FIG. 4 Three-dimensional model results showing acetone distribution, and PAN formation from acetone chemistry. *a*, Acetone distribution with a 49 Tg yr^{-1} total source; *b*, PAN due to acetone chemistry alone. Model results are for 1 March and show a height–latitude cross-section over the central Pacific (175°E longitude).



to 30% in the mid-latitude upper troposphere. As HO_x radicals critically control the production of O_3 via the $\text{NO} + \text{HO}_2 \rightarrow \text{NO}_2 + \text{OH}$ reaction, the presence of acetone in the upper troposphere enhances the production of O_3 . This enhancement may have a number of important implications, in particular for assessing the atmospheric effects of NO_x emissions from subsonic aircraft. Recent HO_x measurements in the lower stratosphere suggest that HO_x sources in addition to those predicted by O_3 photolysis are required to explain the observations¹⁵.

Although PAN is a purely photochemical product which is not emitted directly, sources of oxygenated hydrocarbons have not been subject to rigorous studies. In studies of biogenic emissions from plants, acetone, acetaldehyde, methanol and ethanol have been frequently detected although their emissions have rarely been quantified^{16,17}. Emissions of methanol comparable to those of isoprene have been reported from a variety of plant species¹⁸. In a field study at a forested rural site in the southeastern United States, average mixing ratios of acetaldehyde, acetone, methanol and ethanol of respectively 0.8 p.p.b., 3.3 p.p.b., 6.0 p.p.b. and 1.0 p.p.b. were measured, and a strong biogenic input inferred⁷. Emission of copious quantities of highly reactive methyl butenol $\{\text{C}-\text{C}(\text{C})(\text{OH})-\text{C}=\text{C} \text{ or } \text{C}_5\text{H}_{10}\text{O}\}$, a C_5 alcohol, from vegetation at a forested site in Colorado has also been reported¹⁹. The possibilities that carbonyls and alcohols may be formed from reaction of known biogenic chemicals and water within the plant enzyme system ($\text{C}_5\text{H}_8 + \text{H}_2\text{O} \leftrightarrow \text{C}_5\text{H}_{10}\text{O} \leftrightarrow (\text{CH}_3)_2\text{CO} + \text{C}_2\text{H}_4$; $\text{C}_2\text{H}_4 + 2\text{H}_2\text{O} \leftrightarrow 2\text{CH}_3\text{OH}$; $\text{C}_2\text{H}_4 + \text{H}_2\text{O} \leftrightarrow \text{C}_2\text{H}_5\text{OH}$), and that many of these chemicals may have common precursors, should be explored.

Anthropogenic emissions are also evident from high concentrations of oxygenates that have been reported from polluted areas⁶. In our airborne study, a plume of Asian emissions (China/Hong Kong) was intersected over the western Pacific. Molar ratios of enhancement of species X relative to CO ($\Delta\text{X}/\Delta\text{CO}$) were found to be 0.003 for acetone, 0.007 for methanol and 0.0004 for ethanol. Scaling to a total global CO anthropogenic emission²⁰ of 445 Tg yr^{-1} implies a source of about 2.7 Tg yr^{-1} acetone, 3.8 Tg yr^{-1} methanol and 0.4 Tg yr^{-1} ethanol. There is little doubt that biomass burning is also a source for these species. In a laboratory study²¹ designed to simulate the smouldering combustion of ponderosa pine sapwood, methanol was identified as a product with a molar ratio with respect to CO of 0.025. With global CO biomass burning emissions estimated at 200 Tg yr^{-1} (ref. 20), a methanol source of about 6 Tg yr^{-1} is possible. We have earlier reported that biomass burning

emissions of acetone may be in the vicinity of 10 Tg yr^{-1} (ref. 9).

Oxygenated species are also known to be intermediate products of hydrocarbon oxidation, with methane as the most important source for methanol ($2\text{CH}_3\text{O}_2 \rightarrow \text{CH}_3\text{OH} + \text{CH}_2\text{O} + \text{O}_2$) and propane for acetone^{2,22}. We have attempted to simulate these reactions in our model, after Madronich and Calvert^{14,23}. Assuming removal of methanol by OH oxidation, and by rainout/washout and deposition (0.1 cm s^{-1}) processes, our model calculations (for 1 March) suggest that ~ 160 p.p.t. (global average) of methanol could be synthesized from these reactions, and these mixing ratios may approach 500 p.p.t. in the tropical lower troposphere. Similarly, 100–300 p.p.t. of acetone and 1–10 p.p.t. of ethanol could be synthesized in the middle troposphere via these oxidation mechanisms alone^{9,14,22}.

Yet another source may be found in the oceans. Evidence for supersaturation of some carbonyls in sea water has been published, but the data are sparse and inconclusive²⁴. Oceans are known to be supersaturated with methyl halides²⁵. A possible source for methanol in sea water is from the hydrolysis of methyl halides ($\text{CH}_3\text{X} + \text{H}_2\text{O} \rightarrow \text{CH}_3\text{OH} + \text{HX}$) which takes place at a relatively slow rate²⁶. Suffice it to say that substantial anthropogenic and natural sources of these (and other) oxygenated species are present but remain to be rigorously studied and quantified.

In recent years ethanol has been added to gasoline to achieve cleaner combustion, and the use of ethanol and methanol as biofuels is likely to increase in the future. Measurements of these species in the atmosphere have received attention only recently, in large part because these vapours are difficult to measure, are poorly retained by solid sorbents, and can be easily lost in canisters which are often used as a means of sample collection and storage. Here we have shown that these oxygenated species are globally ubiquitous, can participate in photochemistry of the troposphere in an important way, and have complex biogenic (as well as man-made) sources which require better quantification and further study. It is likely that other oxygenated species are present but have not yet been identified. □

Received 21 April; accepted 19 September 1995.

- Lloyd, A. C. *NBS spec. publ.* **557**, 27–48 (1979).
- Singh, H. B. & Hanst, P. L. *Geophys. Res. Lett.* **8**, 941–944 (1981).
- Singh, H. B. & Zimmerman, P. B. *Adv. env. Sci. Technol.* **24**, 177–235 (1992).
- Blake, D. R. et al. *J. geophys. Res.* **99**, 1699–1719 (1994).
- Cavanagh, L., Schadt, C. & Robinson, E. *Envir. Sci. Technol.* **3**, 251–257 (1969).
- Snider, J. R. & Dawson, G. A. *J. geophys. Res.* **90**, 3797–3805 (1985).
- Goldan, P. D., Kuster, W. C., Fehsenfeld, F. C. & Montzka, S. A. *J. geophys. Res.* (in the press).
- O'Hara, D. & Singh, H. B. *Atmos. Envir.* **22**, 2613–2615 (1988).

9. Singh, H. B. *et al.* *J. geophys. Res.* **99**, 1805–1819 (1994).
10. Arnold, F., Knop, G. & Zeiereis, H. *Nature* **321**, 505–507 (1986).
11. DeMore, W. B. *et al.* *Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling* (JPL Publ. No. 94–26, Jet Propulsion Laboratory, Pasadena, CA, 1994).
12. Betterton, E. A. *Adv. envir. Sci. Technol.* **24**, 1–50 (1992).
13. Atkinson, R. *Atmos. Envir.* **24A**, 1–41 (1990).
14. Kanakidou, M. & Crutzen, P. J. *Chemosphere* **26**, 787–801 (1993).
15. Wennberg, P. *et al.* *Science* **266**, 398–404 (1994).
16. Isidorov, V. A., Zenkench, I. G. & Iofe, B. V. *Atmos. Envir.* **19**, 1–8 (1985).
17. Isidorov, V. A. *Organic Chemistry of the Earth's Atmosphere* (Springer, Berlin, 1990).
18. MacDonald, R. C. & Fall, R. *Atmos. Envir.* **27A**, 1709–1713 (1993).
19. Goldan, P. D., Kuster, W. C., Fehsenfeld, F. C. & Montzka, S. A. *Geophys. Res. Lett.* **20**, 1039–1042 (1993).

20. Lobert, J. M. *et al.* in *Global Biomass Burning* (ed. Levine, J. S.) 289–304 (MIT Press, Cambridge, MA, 1991).
21. McKenzie, L. M., Hao, W. M., Richards, G. N. & Ward, D. E. *Atmos. Envir.* **28**, 3285–3292 (1994).
22. Kanakidou, M., Singh, H. B., Valentin, K. M. & Crutzen, P. J. *J. geophys. Res.* **96**, 15395–15413 (1991).
23. Madronich, S. & Calvert, J. G. *J. geophys. Res.* **95**, 5697–5715 (1990).
24. Zhou, X. & Mopper, K. *Envir. Sci. Technol.* **24**, 1864–1869 (1990).
25. Singh, H. B., Salas, L. J. & Stiles, R. E. *J. geophys. Res.* **88**, 3684–3690 (1983).
26. Elliott, S. & Rowland, F. S. *J. atmos. Chem.* **20**, 229–236 (1995).

ACKNOWLEDGEMENTS. We thank all PEM-West (B) participants for their support. Suggestions by A. R. Ravishankara of the NOAA Aeronomy Lab were appreciated. This work was supported by the NASA Global Tropospheric Experiment, CNRS and CEA.

Experimental evidence for the origin of lead enrichment in convergent-margin magmas

James M. Brenan*, Henry F. Shaw*
& Frederick J. Ryerson*†

* Earth Sciences Division, † Institute for Geophysics and Planetary Physics, Lawrence Livermore National Laboratory, PO Box 808, L-202, Livermore, California 94551, USA

It has been proposed^{1–5} that the low Ce/Pb ratio of subduction-related basalts, relative to their oceanic counterparts, arises by the preferential transfer of lead to the mantle wedge (overlying the subducting slab) by non-magmatic processes. Fluxing of the mantle wedge by low-Ce/Pb fluids, generated by the dehydration of subducted oceanic crust, is one mechanism favoured for this process (see, for example, ref. 5). Here we report the results of a series of high-pressure experiments, which confirm that low-Ce/Pb fluids coexist with the dominant mineral phases (garnet and clinopyroxene) produced during high-pressure dehydration of altered basalt. Our results show that the production of subduction-zone magmas from mantle sources fluxed by basalt-derived fluid is a mechanism by which relatively lead-rich, cerium-poor, mantle-derived material is added to the continents. The lead enrichment of the Earth's continental crust is thus a continuing process occurring at convergent margins.

The Ce/Pb ratio in island-arc basalts (~3) is similar to that of average continental crust, but considerably below that of oceanic basalts (~25) and the bulk Earth (~10; summarized in ref. 5). Owing to similar bulk solid/liquid partition coefficients, these elements are not significantly fractionated during mantle melting processes, and melt compositions reflect the approximate Ce/Pb ratio of their source regions¹. The above-described differences, together with the general similarity between the Ce abundances of island-arc basalts (IABs) and oceanic basalts, suggest that

the source regions for IABs are enriched in Pb relative to the oceanic mantle. As subduction-related magmatism is a continuing process that adds low-Ce/Pb material to the continents, knowledge of the mechanism(s) of Pb-enrichment in IABs may provide insight into both the origin of Pb-enrichment in the continental crust, and the complementary depletion of this element in the oceanic mantle.

Recently, Miller *et al.*⁵ showed that Pb from two isotopically distinct sources is present in the IABs erupted on Unimak island in the Aleutian archipelago. One source, with high ²⁰⁷Pb/²⁰⁴Pb and low Ce/Pb, is consistent with pelagic sediment. The other source, despite having low and mantle-like ²⁰⁷Pb/²⁰⁴Pb, has distinctly non-mantle Ce/Pb (that is, <5 compared to ~25 for the oceanic mantle). Miller *et al.* hypothesized that a low-Ce/Pb mantle component could be produced by the preferential extraction of Pb from subducted oceanic crust by a fluid phase. Transfer of Pb from subducted oceanic crust to arc magmas by this process constitutes a means to explain the origin of the complementary nature of the Ce/Pb ratios of continental crust and oceanic mantle. To test this hypothesis, we have conducted piston-cylinder experiments at 2.0 GPa and 900 °C to determine the Ce/Pb ratio of fluids coexisting with phases that may be residual during high-pressure basalt dehydration, namely clinopyroxene and garnet.

Garnet/fluid and clinopyroxene/fluid partition coefficients for Pb, Ce, Nd, Sm and Yb were measured in experiments that consisted of loading known quantities of trace-element dopant (dried nitrate solutions) with garnet or clinopyroxene and fluid (H₂O ± NaCl) + excess SiO₂ + albite (~3 wt%) in welded Pt capsules. Samples were equilibrated in the piston cylinder for three days, during which time the garnet and clinopyroxene recrystallized: the inclusion of SiO₂ and albite along with the inherent solubility of clinopyroxene and garnet produced a realistic 'eclogite-fluid' composition⁶. After an experiment, capsules were cleaned in concentrated nitric acid, then pierced in ultra-clean distilled water. Newly-formed crystals (pyrope-rich garnet and diopsidic clinopyroxene; see ref. 6) were typically greater than 300 µm in size and were easily removed by hand-picking. The post-experiment fluid, including quenched solids ('fish roe'

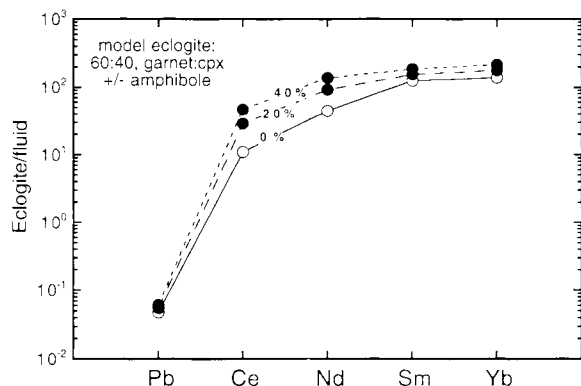


FIG. 1 Bulk eclogite/aqueous fluid partition coefficients for the REE and Pb. Bulk partition coefficients were calculated using the average clinopyroxene and garnet/aqueous fluid values measured in the study and assuming a model eclogite mode (wt% of each mineral in the mixture) of 60:40; garnet:clinopyroxene. Clinopyroxene/aqueous fluid partition coefficients for Ce were determined from the relation $D_{\text{Ce}} = 0.3D_{\text{Nd}}$ (ref. 11) and are consistent with the Ce content of run-product crystals and mass balance (see text). The effect of adding amphibole (20 and 40%) to the residue (as a model for incomplete dehydration) was assessed by calculating amphibole/aqueous fluid partition coefficients assuming values of $D_{\text{REE}}^{\text{amphibole/clinopyroxene}}$ of 3.7 (Ce), 3.1 (Nd), 1.9 (Sm) and 2.5 (Yb), as obtained from mineral/melt partitioning experiments^{11,19}. The amphibole/aqueous fluid partition coefficient for Pb is from ref. 6. Note that the calculated bulk partition coefficients for the REE are more than 200 times larger than for Pb, and this effect is expected to be largest during the onset of dehydration (when the amphibole mode is high).

TABLE 1 Summary of mineral/fluid partition coefficients

Experiment number*	ΣREE	Ce§	Nd		Sm		Yb		Pb
			Rim/fluid	Rayleigh¶	Rim/fluid	Rayleigh	Rim/fluid	Rayleigh	
REE1 cpx	18.9 p.p.m.	[32.0]	106.8 (21.4)	111.1 (11.1)	79.2 (35.0)	118.6 (11.9)	—	—	0.106 (0.012)
REE6 cpx†	2.8 p.p.m.	—	—	84.3 (16.9)	—	108.8 (21.8)	—	138.2 (27.6)	—
REE7 cpx	2.0 p.p.m.	—	—	—	129.0 (29.0)	67.5 (13.5)	—	—	0.171 (0.034)
REE4 cpx‡	16.9 p.p.m.	[23.5]	78.4 (15.7)	67.6 (13.5)	167.3 (33.5)	106.9 (21.4)	122.4 (24.0)	105.4 (21.1)	0.067 (0.013)
REE2 garnet	15.0 p.p.m.	0.34 (0.07)	13.5 (2.7)	—	95.7 (19.1)	—	—	—	0.031 (0.015)
REE8 garnet	1.8 p.p.m.	0.52 (0.10)	15.0 (3.0)	—	126.1 (25.2)	—	145.5 (29.1)	—	0.025 (0.005)

* All fluid compositions were H₂O (except when indicated) with added ~3 wt% silicate material (see text) and contained ~150 p.p.m. Pb with equal amounts of each REE to achieve the indicated total REE concentration (cpx, clinopyroxene).

† Reversal experiment in which REE were initially in presynthesised, REE-doped clinopyroxene powder. Pb was not added to this experiment. Reversed partition coefficients for divalent elements are reported in ref. 6.

‡ Fluid composition was 0.5 m aqueous NaCl.

§ Clinopyroxene/fluid partition coefficients for Ce (shown in square brackets) were calculated from the relation $D_{\text{Ce}} = 0.3D_{\text{Nd}}$ (ref. 11) and agree with measured Ce concentrations in run-product crystals and mass balance. Ce concentrations could not be determined in fluids from these experiments owing to insufficient Ce ionization during IDMS analysis.

¶ Partition coefficients were determined by taking the ratio of rim concentration (typically 2–5 points; measured by SIMS) to the residual fluid concentration (measured by IDMS). Both core and rim concentrations were averaged to determine partition coefficients for Pb. Error on all partition coefficients (1σ , in parentheses) is the larger of either the reproducibility of SIMS analyses or 20% (the estimated uncertainty in the accuracy of the SIMS analyses). Procedural blanks determined for IDMS analysis were negligible compared to concentrations determined in run-products.

* Partition coefficients were determined with a Rayleigh fractionation model using initial and final fluid compositions and the mass fraction of crystals in the experiment.

(small, <20 μm , spheres of silicate glass), orthopyroxene needles), was dissolved in ultrapure nitric and hydrofluoric acids. Trace-element concentrations in run-product crystals were determined by secondary-ion mass spectrometry⁶ (SIMS) and fluid compositions were measured by solid-source isotope dilution mass spectrometry (IDMS) after chemical separation of the rare-earth elements (REEs). Mass balance for the REEs and Pb was monitored in our experiments by comparing the post-experiment inventory for these elements with the total amount added initially. In all cases, mass balance was excellent and no evidence for 'open-system' behaviour (such as loss of Pb to the Pt capsule) was observed. Moreover, reconnaissance SIMS analyses of the Pt capsule retrieved from previous Pb-bearing experiments showed low Pb concentrations and a lack of Pb concentration gradients in the Pt.

Mineral/fluid partition coefficients for the REEs were determined over a concentration range overlapping that of natural basalts (~2–20 p.p.m.), but were kept low enough to prevent the precipitation of REE-oxide phases (compare ref. 7). The partition coefficients were found to be independent of concentration (Table 1). In previous experiments, we have shown this result is also true for Pb partitioning⁶ up to Pb concentrations of ~300 p.p.m. A reversal experiment (REE6), in which all REE were initially present in pre-synthesised, REE-doped clinopyroxene powder, shows excellent agreement with the forward experiments. A demonstration of the reversibility of partition coefficients for divalent cations is provided in ref. 6. These results indicate that Henry's Law is applicable to the behaviour of REEs and Pb during basalt dehydration, and that our partition coefficients represent equilibrium values. We also note that addition of NaCl (0.5 molal) to the fluid did not result in a significant change in partition coefficients for the REE, although values for Pb decreased by up to a factor of ~2.6, which is consistent with previous results⁶. Our clinopyroxene/fluid partition coefficients for the REEs are higher than values previously reported^{8,9}, but subsequent work has shown that these earlier results are probably inaccurate, resulting from either analytical uncertainties¹⁰ and/or precipitation of overlooked REE-rich phases⁷. Owing to the low Ce content of fluids equilibrated with clinopyroxene, insufficient ionization of Ce relative to Nd from these run products precluded reliable determination of their Ce contents. We have therefore estimated the partition coefficient for Ce based on the relationship between the clinopyroxene/melt partition coefficient, $D^{\text{cpx/melt}}$, and REE ionic radius formalized by Blundy and Wood¹¹. Partition coefficients for Ce were derived from

values for Nd by the relation $D_{\text{Ce}} = 0.3D_{\text{Nd}}$, where D_{Ce} and D_{Nd} are the mineral/fluid partition coefficients for Ce and Nd, respectively. Values of D_{Ce} calculated in this manner were found to be consistent with the Ce content of run-product crystals and mass balance.

Assuming that high-pressure dehydration of basalt leaves a residue of garnet and clinopyroxene (60:40 respectively, by mass), bulk rock/fluid partition coefficients, as calculated from average mineral/fluid values, are ~0.05 for Pb and ~11 for Ce (partition coefficients for the other REEs are similarly high; see Table 1 and Fig. 1). For an average mid-ocean-ridge basalt (MORB) having 7.5 p.p.m. Ce and 0.30 p.p.m. Pb (ref. 5), the Ce/Pb ratio of fluid (2 wt%) derived from this material would be ~0.16, which is significantly lower than the initial rock ratio of 25. The presence of residual amphibole enhances the Ce/Pb

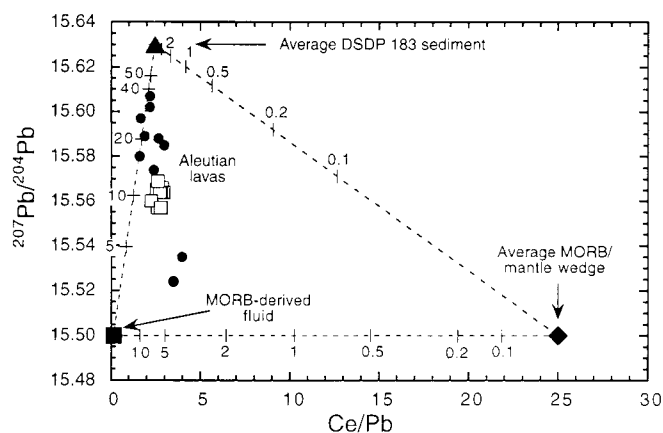


FIG. 2 Lead-isotope composition as a function of Ce/Pb for lavas from Umnak island, Aleutian archipelago (Recheshnoi, open squares; Okmok, filled circles; data from ref. 5), average north Pacific MORB (ref. 5; assumed to be equivalent in Ce/Pb and Pb-isotope composition as a MORB-source-type mantle wedge), average sediment from DSDP Hole 183 (ref. 20, average Ce/Pb from ref. 12) and MORB-derived fluid (Ce/Pb calculated using the data in this study, 2 wt% fluid evolved from a source with a garnet/clinopyroxene mass ratio of 60/40). Numbers on tick marks correspond to the weight percentage of one end-member in each binary mixture. The low- $^{207}\text{Pb}/^{204}\text{Pb}$, low-Ce/Pb component that contributes Pb to the Umnak lavas is consistent with the composition of MORB-derived fluid.

fractionation between solids and coexisting fluids (Fig. 1), as would the presence of NaCl in the fluid. Fluids derived by high-pressure dehydration of material having the isotopic and trace-element composition of MORB will therefore be characterized by Pb-isotope compositions similar to upper-mantle values, but much lower Ce/Pb ratios. Our results confirm the hypothesis of Miller *et al.*⁵ that the low-Ce/Pb, low-²⁰⁷Pb/²⁰⁴Pb component defined by the chemical and isotopic array formed by Umnak island lavas has the characteristics expected for a fluid produced by the high-pressure dehydration of MORB.

The Ce-Pb budgets of lavas from Umnak island are consistent with ternary mixtures of components derived from subducted sediment, slab-derived MORB-fluid, and a mantle wedge with characteristics similar to the MORB source (Fig. 2). By defining the composition of the slab-derived MORB fluid, our new data allow a quantitative assessment of the relative contributions of these components to the source region of Umnak lavas. Calculations assumed that the sediment end-member is simply 'bulk sediment', with average Ce and Pb concentrations of 92 and 38 p.p.m., respectively (values derived by averaging the Ce and Pb concentrations reported in ref. 12) and that slab-derived MORB fluid has 0.7 p.p.m. Ce and 4.5 p.p.m. Pb (derived using the bulk partition coefficients and initial basalt composition given previously). The results show that MORB-derived fluid contributes 60–97 wt% to the slab-derived component (sediment + MORB fluid) and provides 15–70% of the Pb to Umnak lavas. High-pressure dehydration of subducted basalt may therefore be viewed as an important mechanism for generating fluids that reduce the Ce/Pb ratio of the overlying mantle wedge, and thus the source regions for arc magmas. Moreover, subduction of oceanic crust that has lost Pb during dehydration provides a mechanism to increase the Ce/Pb ratio of the oceanic mantle. We conclude that the complementary nature of the crust/oceanic mantle system with regard to Ce/Pb ratios is maintained by a continuing process acting at convergent margins.

We point out, however, that because Archaean subduction zones probably had higher heat flow¹³, direct melting of basalt, as opposed to flux-driven melting of the mantle wedge by basalt-derived fluid, may have been a more important process for generating continental crust early in Earth history (compare ref. 14). Whether this process can produce the requisite fractionation in Ce/Pb is uncertain, owing to the current lack of data regarding the partitioning of these elements during the high-pressure melting of hydrous basalt. But the Archaean tonalite-trondhjemite-granodiorite suite, whose major-element compositions have been reproduced in laboratory experiments on basalt melting^{14–17}, are characterized by low Ce/Pb ratios¹⁸. This suggests that hydrous melting of basalt, and not dehydration, may have generated the Ce/Pb fractionation produced during early crust/mantle differentiation. □

Received 10 July; accepted 25 September 1995.

- Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. *Earth planet. Sci. Lett.* **79**, 33–45 (1986).
- Newsom, H. E., White, W. M., Jochum, J. P. & Hofmann, A. W. *Earth planet. Sci. Lett.* **80**, 299–313 (1986).
- Hofmann, A. W. *Earth planet. Sci. Lett.* **90**, 297–314 (1988).
- Peucker-Ehrenbrink, B., Hofmann, A. W. & Hart, S. R. *Earth planet. Sci. Lett.* **125**, 129–142 (1994).
- Miller, D. M., Goldstein, S. L. & Langmuir, C. H. *Nature* **368**, 514–520 (1994).
- Brenan, J. M., Shaw, H. F., Ryerson, F. J. & Phinney, D. L. *Geochim. cosmochim. Acta* **59**, 3331–3350 (1995).
- Meen, J. K., Eggle, D. H. & Ayers, J. C. *Nature* **340**, 301–303 (1989).
- Mysen, B. O. *Am. Mineral.* **64**, 274–287 (1979).
- Mysen, B. O. *Neues Jb. Miner. Abh.* **146**, 41–65 (1983).
- Beattie, P. *Geochim. cosmochim. Acta* **57**, 47–55 (1993).
- Blundy, J. & Wood, B. *Nature* **372**, 452–454 (1994).
- Ben Othman, D., White, W. M. & Patchett, P. J. *Earth planet. Sci. Lett.* **94**, 1–21 (1989).
- Martin, H. *Geology* **14**, 753–756 (1986).
- Rapp, R. P., Watson, E. B. & Miller, C. F. *Precamb. Res.* **51**, 1–25 (1991).
- Beard, J. S. & Lofgren, G. E. *J. Petrol.* **32**, 365–401 (1991).
- Rushmer, T. *Contr. Miner. Petrol.* **107**, 41–59 (1991).
- Wolf, M. B. & Wyllie, P. J. *Contr. Miner. Petrol.* **115**, 369–383 (1994).
- Weaver, B. L. & Tarney, J. *Earth planet. Sci. Lett.* **51**, 279–296 (1980).

- Brenan, J. M., Shaw, H. F., Ryerson, F. J. & Phinney, D. L. *Earth planet. Sci. Lett.* (in the press).
- McDermott, F. M. & Hawkesworth, C. M. *Earth planet. Sci. Lett.* **104**, 1–15 (1991).

ACKNOWLEDGEMENTS. The manuscript benefited from the comments of C. Chauvel. This work was supported by the US Department of Energy's Office of Basic Energy Sciences and LLNL's branch of the Institute of Geophysics and Planetary Physics. Work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory.

Scaling of elastic strain energy in kangaroos and the benefits of being big

M. B. Bennett & G. C. Taylor

Department of Anatomical Sciences, University of Queensland, St Lucia, Queensland 4072, Australia

LARGE kangaroos are unique among mammals in their ability to uncouple aerobic metabolic energy costs from the speed of locomotion^{1,2}, making hopping an economical gait. During the first half of the ground-contact phase, kinetic energy lost from the body is stored as elastic strain energy, predominantly in the hind limbs^{3–5}. The subsequent recoil returns kinetic and potential energy to the body. Here we show that the allometry of structures in the legs and feet of Macropodoidea is different from that of quadrupedal eutherian mammals. The potential for elastic energy storage in hoppers is shown to scale with strong positive allometry. This is a function of the structural properties of muscle-tendon units in the distal hind limbs and the postures adopted by hopping kangaroos. Our findings demonstrate how the use of tissue elasticity is strongly mass dependent and help explain the observed energetic phenomena.

Kangaroos, wallabies and rat-kangaroos use a variety of gaits when moving slowly, but most hop bipedally at high speeds⁶. 'Slow progression' is accompanied by linear increases in the mass-specific metabolic energy cost with speed, as is the hopping of small macropods^{1,2}. Locomotor costs are similar to those for quadrupedal mammals of comparable body size^{2,7}. However, when large kangaroos change to a hopping gait, the aerobic metabolic cost of locomotion becomes independent of speed, giving them an energetic advantage over eutherian mammals, especially at high hopping speeds^{1,2}. This phenomenon is not completely understood, although tissue elasticity is thought to be important.

The major elastic strain energy stores in kangaroos are likely to be collagenous structures in the hind limbs, which are subjected to large forces during ground contact, especially those which act to balance rotational moments at the ankle⁸. These are tendons of the muscles gastrocnemius, flexor digitorum superficialis (fds) and flexor digitorum profundus (fdp). To assess their potential for strain energy storage, we dissected and measured components of the hind limbs of 26 macropods (mass, 0.5–38.0 kg).

If kangaroos were geometrically similar (where shape remains constant and all linear dimensions scale proportional ($\propto$) to body mass (M)^{0.33}), muscle-fibre and tendon lengths would be $\propto M^{0.33}$, physiological cross-sectional areas of muscles (PCSA_m)⁹ and cross-sectional areas of their attached tendons (CSA_t) would be $\propto M^{0.67}$, and muscle masses would be $\propto M^{1.00}$. Eutherian mammals conform broadly to these predictions, with the exception of muscle-fibre lengths, which scale with negative allometry¹⁰. However, the allometry of comparable structures in kangaroos is quite different for most of the measured parameters (Fig. 1 and Table 1).

The ratio of PCSA_m:CSA_t when multiplied by the maximum isometric muscle stress (taken here to be 300 kPa¹⁰) yields a value for the maximum stress to which a tendon can be exposed.

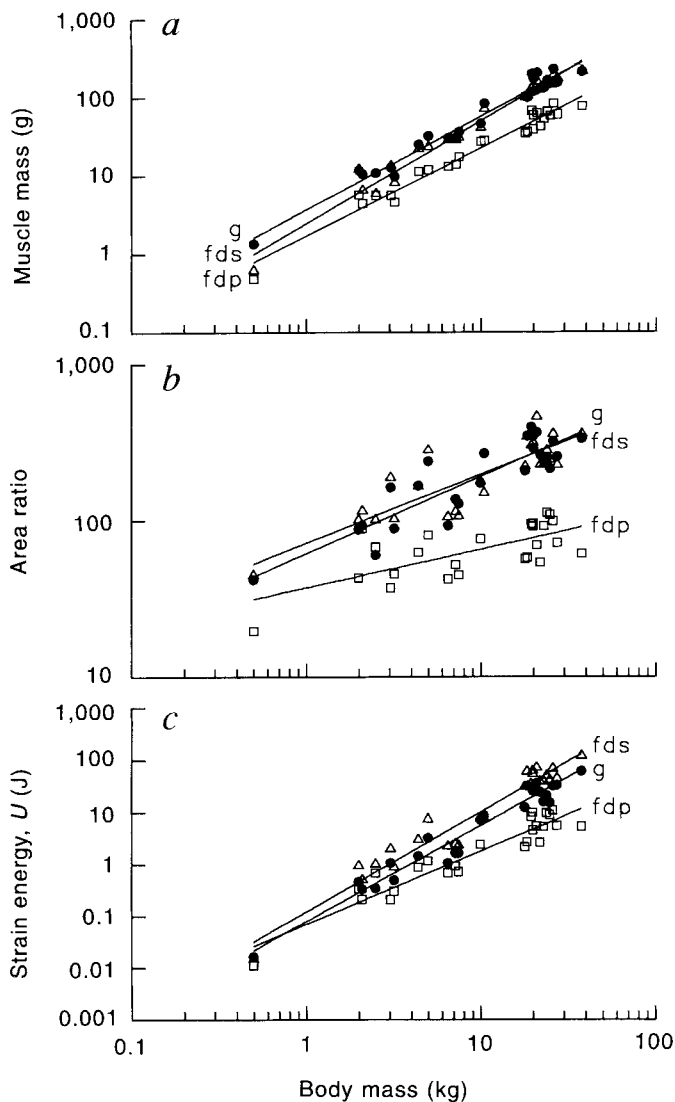


FIG. 1 Graphs of logarithmic coordinates of the relationships between muscle mass (a), ratio of physiological muscle area to tendon area (b), and maximum capacity for elastic strain energy storage (c) and body mass for the major muscles of plantarflexion of the ankle in 10 species of macropod marsupials. Least-squares regression lines are shown for gastrocnemius (g, filled circles), flexor digitorum superficialis (= plantaris) (fds, open triangles) and flexor digitorum profundus (fdp, open squares). Allometric equations describing these data are given in Table 1. Area ratios were not corrected for angles of muscle-fibre pinnation or sarcomere spacing⁹. This allows direct comparisons to be made with results from quadrupedal mammals¹⁰. However, pinnation angles and sarcomere spacings were, respectively, $28.5 \pm 7.5^\circ$ (mean and s.d., $n=52$) and $1.90 \pm 0.24 \mu\text{m}$ ($n=38$) for gastrocnemius; $34.5 \pm 10.6^\circ$ ($n=26$) and $1.91 \pm 0.21 \mu\text{m}$ ($n=19$) for flexor digitorum superficialis; and $31.0 \pm 6.5^\circ$ ($n=26$) and $2.07 \pm 0.17 \mu\text{m}$ ($n=19$) for flexor digitorum profundus. Values for the maximum capacity for strain energy storage were calculated¹⁰ by reference to a tensile stress-strain curve typical of tendons from the eutheria and metatheria^{5,11}. The linear elastic modulus of the tendon is 1.425 GPa. The method of calculation does not allow for the curved toe region of typical stress-strain curves¹¹ and will slightly underestimate elastic strain energy storage. Other structures capable of acting as strain energy storage sites are the focus of continuing investigation. In all cases the exponents describing the regression lines were significantly greater ($P < 0.001$) than predicted by geometric similarity. The species studied were *Macropus rufus*, *M. giganteus*, *M. eugenii*, *M. robustus*, *Thylogale thetis*, *T. stigmatica*, *Aepyprymnus rufescens*, *Potorous tridactylus*, *Hypsiprymnodon moschatus* and *Wallabia bicolor*.

The area ratios and therefore peak tendon stresses show strong positive allometry (Fig. 1 and Table 1). A stress-strain curve for the tensile loading of mammalian tendons¹¹ was used to calculate the potential for elastic strain energy storage (U) in each tendon. The storage capacity for these kangaroo tendons is proportional to $(M)^{1.73}$, whereas for equivalent structures in quadrupedal mammals¹⁰, $U \propto M^{1.28}$. Thus, although both groups

can store about 0.6 J of strain energy at a body mass of 1.6 kg, a 50-kg kangaroo would be able to store 357 J, in contrast to just 55 J for a eutherian mammal of the same size.

Significant energy storage will only occur if large stresses act in these structures during the ground-contact phase. We analysed high-speed films of eight species ranging in size from *Potorous tridactylus* (1 kg) to *Macropus giganteus* (50 kg). Joint

TABLE 1 Allometric equations for morphological and biomechanical characteristics in relation to locomotion in the Macropodoidea

	Gastrocnemius	(r)	Fds	(r)	Fdp	(r)
Muscle mass (g)	$3.69(M)^{1.19 \pm 0.04}$	(0.98)	$2.46(M)^{1.30 \pm 0.04}$	(0.99)	$1.71(M)^{1.12 \pm 0.04}$	(0.98)
Muscle area (mm ²)	$214(M)^{1.06 \pm 0.05}$	(0.98)	$169(M)^{1.19 \pm 0.06}$	(0.99)	$145(M)^{0.81 \pm 0.07}$	(0.93)
Tendon area (mm ²)	$3.52(M)^{0.57 \pm 0.04}$	(0.95)	$2.34(M)^{0.75 \pm 0.04}$	(0.97)	$3.85(M)^{0.56 \pm 0.05}$	(0.93)
Muscle fibre length (mm)	$15.7(M)^{0.14 \pm 0.04}$	(0.62)	$13.9(M)^{0.12 \pm 0.05}$	(0.45)	$11.1(M)^{0.32 \pm 0.04}$	(0.86)
Tendon length (mm)	$76.8(M)^{0.49 \pm 0.02}$	(0.98)	$138(M)^{0.47 \pm 0.02}$	(0.98)	$132(M)^{0.47 \pm 0.02}$	(0.97)
Area ratio	$61.8(M)^{0.49 \pm 0.05}$	(0.88)	$72.1(M)^{0.44 \pm 0.06}$	(0.64)	$37.5(M)^{0.24 \pm 0.06}$	(0.67)
Strain energy capacity (J)	$0.08(M)^{1.85 \pm 0.08}$	(0.98)	$0.12(M)^{1.93 \pm 0.10}$	(0.97)	$0.07(M)^{1.41 \pm 0.10}$	(0.95)
Moment arm, r (mm)	$10.7(M)^{0.46 \pm 0.02}$	(0.97)				
Moment arm, R (mm)	$44.7(M)^{0.97 \pm 0.02}$	(0.97)				
EMA	$0.24(M)^{0.00 \pm 0.02}$	(0.03)				
Duty factor	$0.65(\text{speed})^{-0.39 \pm 0.04}$	(0.88)				

Allometric equations were obtained by least-squares regression of data from 26 kangaroos following logarithmic transformation, and are reported in the form $y = ax^b$. The top part refers to the muscle-tendon units of gastrocnemius, flexor digitorum superficialis and flexor digitorum profundus referred to in the text and in Fig. 1. The bottom refers to Figs 2 and 3. Body mass, M , is in kg, and speed is in m s^{-1} . The correlation coefficients (r) for each regression and standard errors for each exponent (b) are given.

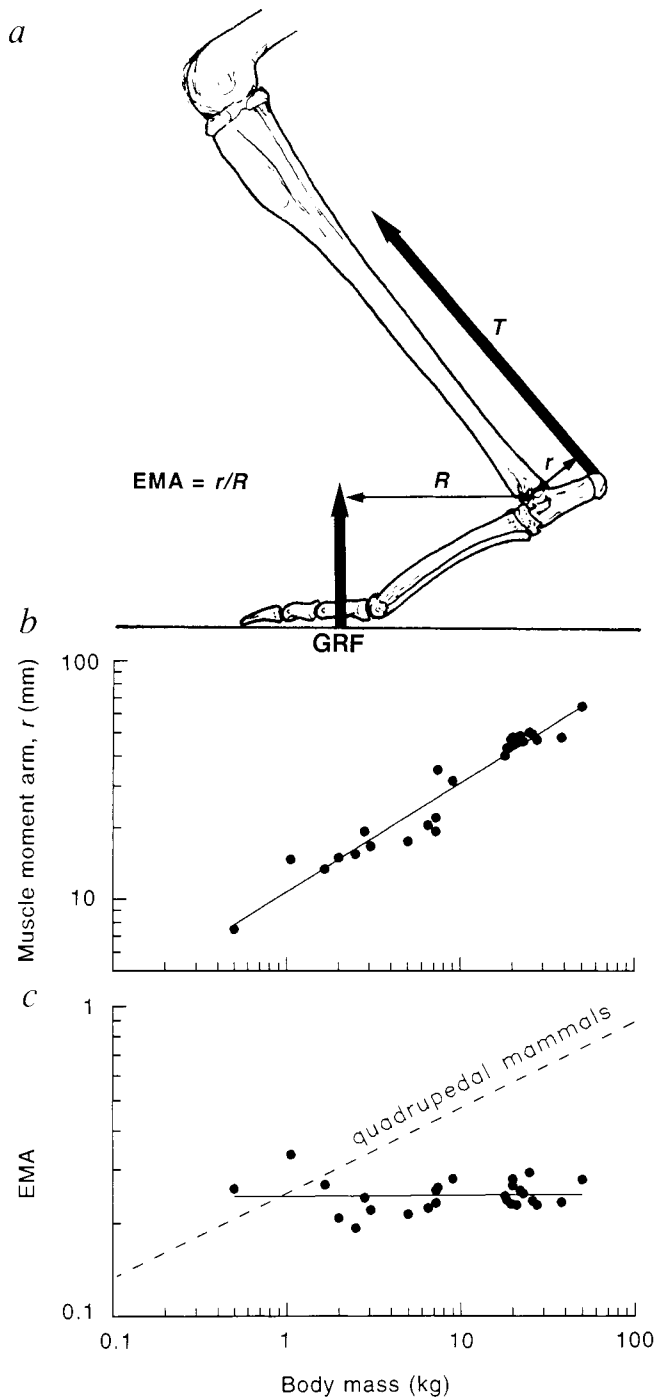


FIG. 2 Diagram of the crus and pes of a western grey kangaroo (*M. giganteus*) taken from a radiograph. The joint angles represent those recorded at mid-stance for macropods hopping overground and on treadmills. Only the large digit IV is shown, the smaller digits I (when present), II, III and V having been omitted for clarity. The ground reaction force, GRF, was conservatively assumed to act at the mid-point of the proximal phalanx of digit IV. The moment arm, R , was taken as the perpendicular distance from the line of action of this force to the joint centre of the ankle. Tension (T) in the tendons of gastrocnemius and fds, acting with a moment arm r , act to balance moments at the ankle (flexor digitorum profundus also assists in balancing these moments, but acts with a moment arm of about $r/3$). *b*, The scaling of the muscle moment arm (r) about the ankle plotted as a function of body mass on logarithmic coordinates. The slope of the least-squares regression line (0.457) is similar to that reported for quadrupedal mammals (0.434) (ref. 12). *c*, Plot on logarithmic axes of muscle effective mechanical advantage, EMA, as a function of body mass. EMA is independent of body mass in the Macropodoidea, because the moment arm for the GRF increases at the same rate as r with body mass. In quadrupedal mammals (dashed line from ref. 12) the EMA is proportional to $M^{0.27}$, reflecting the more upright limb postures adopted by larger quadrupedal mammals. The Macropodoidea need 'crouched-limb' postures for effective hopping locomotion irrespective of body size. Least-squares regression line data are given in Table 1.

angles determined at mid-stance showed hind limb postures to be similar in all species, allowing lines of action of muscles to be estimated (Fig. 2). Exponents describing the moment arms, r of gastrocnemius and fds and R of the peak vertical ground reaction force (GRF), about the ankle were similar. Consequently, the effective mechanical advantage ($EMA = r/R$; Fig. 2) is independent of body mass. This is in marked contrast to the situation in quadrupedal mammals¹², where a more vertical alignment of limb segments in larger animals results in $EMA \propto M^{0.27}$. Kangaroos do not adopt a more straight-limbed stance as they increase in size, presumably because hopping gaits require bent-limb postures to be effective. The mass-independent EMA means that muscle forces needed to balance moments at the ankle increase rapidly with body size.

The fraction of time for which the feet are on the ground (the duty factor) was found to be proportional to $(\text{speed})^{-0.39}$ (Fig.

3). As duty factors approximate to GRF^{-1} (ref. 13) then GRF is proportional to $(\text{speed})^{0.39}$. Thus, at any given speed, average tendon forces will be proportional to $M^{0.39}$. This is consistent with the larger than predicted exponents for PCSA_m of gastrocnemius, fds and fdp, all of which act to resist ankle dorsiflexion. Maximum attainable speeds for kangaroos are correlated with body mass¹⁴ ($u_{\max} \propto M^{0.36}$), suggesting that GRFs for small animals are limited to smaller multiples of body mass than for large ones. This is important for strain energy storage and tendon safety factors (Fig. 3).

At a given speed, plantarflexor tendons are subjected to higher stresses in larger individuals than in smaller ones. Larger kangaroos can also attain higher speeds and so have an even greater potential for strain energy storage. As a first approximation, about 35% (37 J s^{-1}) of the mechanical work rate³ of a 20-kg kangaroo can be accounted for at 2.8 m s^{-1} , rising to 42%

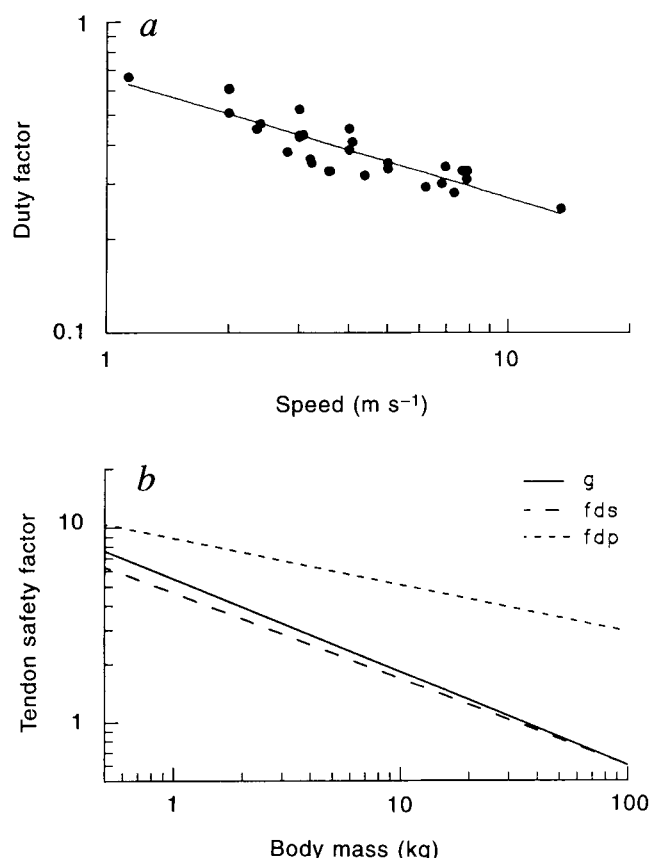


FIG. 3 Graphs on logarithmic coordinates of duty factor versus speed of hopping (a) and tendon safety factors versus body mass (b) in the Macropodoidea. Duty factors were determined from analysis of high-speed cine and video films of 8 species. Individual animals ranged from 1 kg to 50 kg body mass. Data were obtained from previous studies^{3,4,14,17} and unpublished observations. Film footage included sequences of some species hopping at or close to their maximum speeds. Regression information is given in Table 1. Tendon safety factors (tendon strength/maximum stress to which the tendon is likely to be exposed) were calculated assuming a peak isometric muscle stress of 300 kPa. The ultimate tendon strength was taken to be 100 MPa¹¹. Tendons with low safety factors make good 'springs' for hopping, but have a greater risk of failing under the applied loads. Least-squares regression line data are: $SF_{\text{gastrocnemius}} = 5.4M^{-0.49 \pm 0.05}$; $SF_{\text{fds}} = 4.6M^{-0.44 \pm 0.06}$; $SF_{\text{fdp}} = 5.4M^{-0.24 \pm 0.06}$.

(97 J s⁻¹) at 6.8 m s⁻¹. These are lowest estimates based on tendon forces being proportional to muscle-fibre areas⁵. Greater savings would be expected in larger, faster animals, and also if the pattern of muscle force sharing¹⁵ was different.

A continuum exists from small macropods in which energy savings are trivial to large ones where tendon elasticity significantly reduces the cost of travel. However, 50–60 kg may be the optimum mass for this energy-saving system as larger animals risk breaking their tendons (Fig. 3). Extinct giant kangaroos of the Pleistocene (~150 kg)¹⁶ probably used lower than predicted hopping speeds or may have scaled differently to modern species. In either case, the potential benefits of elastic energy storage would have been reduced, perhaps contributing to their eventual extinction. □

Received 30 May; accepted 4 September 1995.

1. Dawson, T. J. & Taylor, C. R. *Nature* **246**, 313–314 (1973).
2. Baudinette, R. V., Snyder, G. K. & Frappell, P. B. *Am. J. Physiol.* **262**, R771–R778 (1992).
3. Cavagna, G. A., Heglund, N. C. & Taylor, C. R. *Am. J. Physiol.* **233**, R243–R261 (1977).
4. Alexander, R. McN. & Vernon, A. J. *Zool., Lond.* **177**, 265–300 (1975).
5. Ker, R. F., Dimery, N. J. & Alexander, R. McN. *J. Zool., Lond. (A)* **208**, 417–428 (1986).
6. Windsor, D. E. & Dagg, A. I. *J. Zool., Lond.* **163**, 165–175 (1971).
7. Thompson, S. D., MacMillan, R. E., Burke, E. M. & Taylor, C. R. *Nature* **287**, 223–224 (1980).

8. Alexander, R. McN. *Elastic Mechanisms in Animal Movement* (Cambridge Univ. Press, 1988).
9. Bennett, M. B. *J. Zool., Lond.* **235**, 33–42 (1995).
10. Pollock, C. M. & Shadwick, R. E. *Am. J. Physiol.* **266**, R1022–R1031 (1994).
11. Bennett, M. B., Dimery, N. J., Ker, R. F. & Alexander, R. McN. *J. Zool., Lond.* **209**, 537–548 (1986).
12. Biewener, A. A. *Science* **250**, 1097–1103 (1990).
13. Alexander, R. McN., Maloy, G. M. O., Njau, R. & Jayes, A. S. *J. Zool., Lond.* **187**, 169–178 (1979).
14. Bennett, M. B. *J. Zool. Lond.* **212**, 457–464 (1987).
15. Herzog, W., Leonard, T. R. & Guimaraes, A. C. S. *J. Biomech.* **26**, 945–953 (1993).
16. Murray, P. in *Vertebrate Palaeontology of Australasia* (eds Vickers-Rich, P., Monaghan, J. M., Baird, R. F. & Rich, T. H.) 1071–1164 (Monash Univ. Publications Committee, Melbourne, 1991).
17. Farley, C. T., Glasheen, J. & McMahon, T. A. *J. exp. Biol.* **185**, 71–86 (1993).

ACKNOWLEDGEMENTS. We thank R. V. Baudinette and A. A. Biewener for film of hopping kangaroos.

Genetic polymorphism for alternative mating behaviour in lekking male ruff *Philomachus pugnax*

David B. Lank, Constance M. Smith, Olivier Hanotte*, Terry Burke* & Fred Cooke

Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

* Department of Zoology, University of Leicester, Leicester LE1 7RH, UK

ALTERNATIVE male mating tactics are widespread among animal taxa^{1–3}, but there are few well documented examples of genetic polymorphisms for them^{4–6}. The dimorphism in male courtship behaviour between independent and satellite ruffs, *Philomachus pugnax*^{7,8} (a lekking sandpiper), has often been cited as a potential example but this has been questioned^{9,10} because of the lack of data¹¹ and the widespread phenotypic plasticity in the development or expression of alternative tactics in other species^{1–3,9,12–14}. By rearing ruffs in captivity, we now show that differential morph development is genetically controlled and consistent with a single-locus, two-allele autosomal genetic polymorphism. Several potentially relevant environmental factors do not appear to alter behavioural development.

'Independent' male ruffs have a territorial breeding strategy of defending lek mating courts against other independents. Non-territorial 'satellites', representing about 16% of males¹⁵ (D.B.L. and C.M.S., unpublished data), move among, are recruited to and share independents' courts^{7,8,11}. Differences in mating strategy correlate with polymorphisms in the colours and patterns of elaborate male breeding plumage⁷. Plumage variation is similar within populations throughout the species' Palaearctic breeding range, indicating that behavioural variation is also ubiquitous. Males maintain a single plumage and behavioural phenotype throughout their adult lifetime. A single observation of short-term behavioural flexibility in the wild⁸ is counterbalanced by a failure to report another case in many years of fieldwork by six research groups, including the original observer^{7,11,16–19}.

We obtained pedigree data for behavioural morph in two ways. In 1985, 1989 and 1990, we reared chicks from eggs collected from 43 ruff nests at three sites near Oulu, Finland. The paternity of the 1989 and 1990 chicks was determined by comparing a chick's DNA profile with that of its mother and of males captured at nearby leks (Fig. 1). From 1987 to 1993, we bred the ruffs we reared, and their offspring, in captivity, controlling paternity by housing females with a single male.

Table 1 presents pedigree data for behavioural phenotype by family. Of 27 sons of independent males, 25 were independents, whereas 14 of 30 satellite sons were satellites (corrected $\chi^2 =$

TABLE 1 Pedigree data on the behavioural phenotype of ruff sons, their fathers, and their mother's brothers, classified by their father's and their maternal grandfather's phenotype

Source*	Father	Mother	Mother's brother's morph(s)†‡	Morph(s) of son(s)†
Independent fathers × independent maternal grandfathers				
C	112	28	—	I, I
C	112	37	—	I
C	114	40	I	I, I
C	151	73	—	S
Independent fathers × unknown maternal grandfathers				
C§	G1M	G1F	—	I
W	B&W	K5	—	I
W	B&W	K10	—	I
W	SS	L11	—	I
W	LO	L13	—	I
W	LO	L15	—	I, I
C	112	03	I	I
C	112	05	I	I
C	112	11	I	I
C	112	18	I	I, I
C	112	U1	—	I
C	114	07	I, I	I
C	6247	11	I	I
C	6247	U2	—	I
C	6247	U3	—	I, I, I
Independent fathers × satellite maternal grandfathers				
C	112	67	—	S
C	113	24	I	I, I
C	114	62	I	I
Satellite father × independent maternal grandfathers				
C	111	59	I, I	S, I
C	111	53	—	S
Satellite fathers × unknown maternal grandfathers				
C§	G2M	G2F	—	S
W	BT	K19	—	I
W	BT	K22	—	I
W	TH	K3	—	I
W	BU	L5	—	I
W	BU	L16	—	I
W	BF	L16	—	I
C	111	11	I	I
C	111	17	I	S, S, I, I, I
C	111	61	—	S
C	158	6174	—	S, I
Satellite fathers × satellite maternal grandfathers				
C	111	22	S, S, I, I, I	S
C	111	71	S	S
C	111	6233	—	S, I, I
C	118	24	I	I
C	158	22	S, S, I, I, I	S
C	158	24	I	S, S, I
C	158	510	S, I	S

Chicks were reared in 'common garden' social groups that mixed individuals from different broods, with no adults present through the first year (1985 and 1989 cohorts) or the first 2 months of life (other years). Males of both morphs developed, even within the two cohorts raised in the absence of adults. We assessed the behavioural phenotypes of males by observing male–male interactions in mixed-morph groups during breeding seasons. The relationship between plumage and behavioural phenotype seen in nature⁷ also developed in captivity. No birds changed morphs among years. We biased breeding assignments towards positive assortative mating to enhance our probability of detecting genetic differences, using expectations about a female's potential genotype based on the phenotype(s) of her male relatives, where known.

* W, DNA paternity analysis of wild-bred chicks reared in captivity (Fig. 1); C, bred in captive flock from birds previously raised.

† I, Independent; S, satellite.

‡ Brothers of females with unknown grandfathers may be half or full sibs.

§ Published data¹⁵.

|| 1–3 females' offspring.

¶ Pedigree class assigned from a satellite brother, which implies a satellite grandparent under a dominant satellite allele model (inclusion with 'unknown grandfather' does not change the outcome of the analyses).

Note, Birds 22, 24, 118 and 510 are offspring of 111; 118 is a son of 17; 28, 37 and 53 are daughters of 114; 59 and 73 are daughters of LO; 40 is a daughter of 112; 6233 is a daughter of TH; 67 is a daughter of BU; 62 is a daughter of BF, and 510 is a daughter of 59.

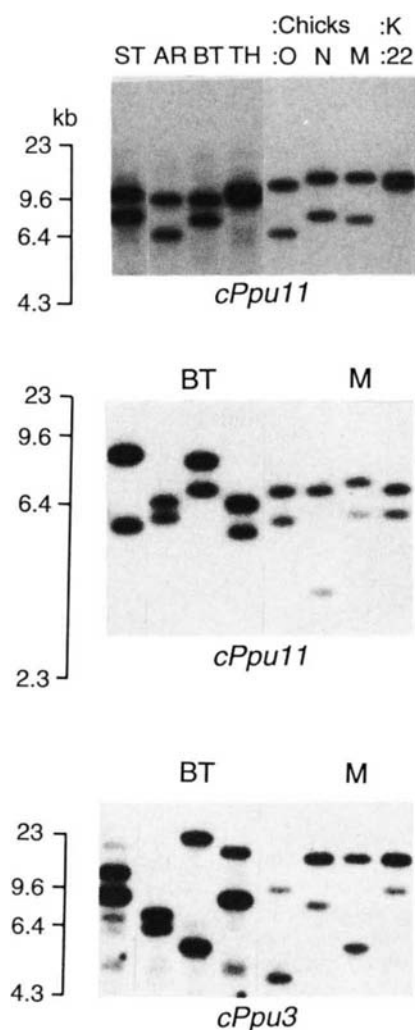


FIG. 1 Example of paternity assignment of a wild-caught chick using single-locus minisatellite probes. The paternal alleles of chick M, from female K22's nest, are only present in satellite father BT. Five single-locus minisatellite probes (*cPpuMS3*, *cPpuMS4*, *cPpuMS6*, *cPpuMS8*, *cPpuMS11*) were isolated from a ruff size-selected genomic DNA library following a standard procedure^{27,28}. The minisatellite loci were characterized by a large number of alleles (12.6 ± 1.82 s.d.) in our population. The probes were hybridized to a Southern blot of total *Mbol*-restricted genomic DNA from the chicks, their biological mothers and potential fathers. Males were run on the same gel (chicks from females K3, K5, K10, K19, K22) or on a different gel (chicks from females L5, L11, L15, L16) from the chicks. In the latter case, an internal size marker mix was added to each digested DNA sample²⁹. At least four probes were used for paternity assignments. Paternities were assigned only when an exact size matching of the paternal allele for all probes was achieved. Assuming an equal allele frequency at each locus, the paternal false inclusion probability was estimated to be $<5.4 \times 10^{-4}$ for each combination of four probes used (ref. 30, and D.B.L. et al., manuscript in preparation).

29.87, 1 d.f., $P=0.002$), indicating a strong effect of paternal inheritance on offspring phenotype. We tested observed offspring morph ratios against the predictions from four single-locus, two-allele genetic models with complete dominance, two each for autosomal and sex-linked (Z chromosome) genetic inheritance, assuming Hardy–Weinberg equilibrium in the grandparent generation (Table 2). Because females do not express the behavioural phenotype, we classified families by father's and maternal grandfather's phenotypes, including unknown grandparent cases, producing six pedigree classes (Tables 1, 2a). Offspring morph ratios were heterogeneous among classes (ANOVA of offspring ratios per male, weighted

TABLE 2 'Goodness of fit' tests of autosomal models for the inheritance of male mating strategies

(a) Morphs produced under satellite (Ss) and independent (Ii) complete dominance autosomal models

Pedigree*		n‡	Proportion of independent male offspring			P(Ss)	Combined	P(Ii)	Combined
Father	MGF†		Expected (Ss)	Expected (Ii)	Observed				
I	I	6	0.96	0.92	0.83	0.23	0.46	0.46	0.74
I	U	19	0.92	0.89	1.00	0.40		0.16	
I	S	4	0.70	0.80	0.75	1.00		0.59	
S	I+U§	19	0.44	0.61	0.62	0.11		1.00	
S	S	11	0.33	0.30	0.36	0.76		0.74	

(b) Heterozygote phenotype under both models

Phenotype of father	Genotype	Expected frequencies	Proportion of crosses producing both phenotypes			Goodness of fit	
			Expected	Observed	n_{ij}	P	Combined P
Satellite dominance model							
Independent	<i>ss</i>	0.84	0.08	0.00	5	1.00	0.49
Satellite	<i>Ss</i>	0.15	0.67	1.00	5	0.18	
Satellite	<i>SS</i>	0.01					
Independent dominance model							
Independent	<i>Ii</i>	0.36	0.15	0.00	5	1.00	0.03
Independent	<i>Ii</i>	0.48					
Satellite	<i>ii</i>	0.16	0.35	1.00	5	0.01	

a, We compared the observed proportions of independent male offspring produced by residents and satellites crossed with three classes of maternal grandfather, including unknown morph, against expected values generated under each model. Expected allelic and genotypic frequencies in the grandparent generation were calculated from phenotypic frequencies of 0.84 independents and 0.16 satellites observed in the wild, assuming Hardy-Weinberg equilibrium, producing $p(S) = 0.08$, $q(s) = 0.92$ for the satellite dominance model, and $p(I) = 0.60$, $q(i) = 0.40$, for the independent dominance model, producing the expected genotype frequencies shown in b. Assortative mating would be one source of departure from Hardy-Weinberg equilibrium. Because the eggs of individual females are often fertilized by males of both morphs in the wild, this does not strongly occur (D.B.L. *et al.* unpublished data). The goodness of fit was tested using 2-tailed binomial probabilities, and an overall probability was calculated using Fisher's method for combining probabilities, d.f. = 4. Neither model is rejected for any single pedigree class (2-tailed binomial test), nor for the data set as a whole. Because different numbers of offspring per parent were used in this analysis, the genotype ratios within the dominant allelic class may not be representative of the population as a whole, and the probability of mistakenly rejecting a valid model increases. Our failure to reject either model despite this risk is thus conservative. b, We tested for the heterozygosity of independent versus satellite males by comparing the proportion of families with multiple male offspring of both morph types (Table 1) against expected proportions. The expected values were calculated under Hardy-Weinberg equilibrium and incorporated genotype-specific probabilities for producing mixed families given the number of males produced. Because we attempted to assort captive matings positively, which would bias against the production of mixed-morph families, the probabilities calculated for rejection of both models are conservative. The independent dominance model is rejected, because of a poor fit for satellites, using Fisher's method of combining probabilities, d.f. = 1.

*I, Independent; S, satellite; U, unknown.

† Maternal grandfather.

‡ n, Number of offspring per cross.

§ Because data from only three offspring were obtained from the satellite × independent pedigree (Table 1), and the expected morph ratios were most similar to those from pedigrees of unknown grandfathers, these two classes were combined and tested against sample-size weighted expected values.

|| n, Number of families.

by number of offspring: $n = 24$, $F = 3.68$, 5 d.f., $P = 0.018$), further demonstrating the effect of pedigree.

Neither sex-linked complete dominance model fits the data. Male birds are homogametic (ZZ), so neither model allows for the production of opposite phenotype sons from both matched-morph father × maternal grandfather pedigrees, as occurs in Table 1. Rejection of the sex-linked satellite dominance model relies on an assignment of paternal pedigree for female 73 based on DNA fingerprinting (Fig. 1)

TABLE 3 Gompertz growth parameters²⁶ and body mass of independent and satellite males as chicks

Variable	Independents			Satellites			F	P
	mean	s.e.	n	mean	s.e.	n		
K	0.234	0.004	59	0.245	0.009	17	1.57	0.22
t	10.3	0.2	59	9.5	0.4	17	3.40	0.07
a	142.5	1.8	59	134.2	4.0	17	3.62	0.06
Adult mass	167.3	1.9	57	155.9	4.7	14	5.14	0.003

Satellites are smaller, but growth parameters (K and t) do not differ as expected if poorly growing chicks develop into satellites. Mass of growing chicks was measured on alternate days through to day 26. K, Maximal slope of growth curve (1/day); t, inflection point of growth curve (days); a, asymptotic body mass at the end of chick growth (g); adult mass is the mass or mean mass measured during mid-winter(s) as adults (g). Variable values are least-square means from ANCOVA models controlling for significant differences in cohort growth rate; F and P values test for differences between phenotypes (1 d.f.).

Observed offspring morph ratios produced by each pedigree class fit expectations from both autosomal dominance models (Table 2a). We may discriminate between the two by specifically testing their predictions about the frequencies and phenotypes of heterozygotes. The satellite allele dominance model predicts that 15% of all males will be heterozygotes and satellites, whereas the independent dominance model predicts that 48% of males will be heterozygotes and independents (Table 2). Only crosses in which at least one parent is a heterozygote can produce mixed broods. Among families that produced multiple sons, all five satellite crosses produced both morphs, whereas all five independent crosses produced only independents (Table 1). The independent dominance model is rejected by this outcome (Table 2b), given the probabilities of mating with heterozygous females, assuming Hardy-Weinberg equilibrium, and the family sizes involved. More complex modes of inheritance cannot yet be adequately tested, but the satellite dominance model appears to account for the inheritance of behavioural phenotype.

Because our results pertain strictly to male development under captive conditions, natural environmental variation, eliminated in our study, could significantly affect phenotype development or expression^{1, 3, 8, 14}. Captive rearing did not drastically bias development, because 11% (5 of 45) of the males reared from eggs collected in the wild became satellites, similar to the proportion in the source population (16%). We eliminated common

environmental effects as a cause of father-son phenotypic similarity by rearing young in groups (Table 1).

A slightly smaller mean body size of wild satellites has been interpreted as suggesting that poorly growing chicks disproportionately develop into satellites¹⁰. Small body size differences between morphs developed among our captives, despite *ad libitum* food availability and benign environmental conditions (Table 3). However, growth parameters do not indicate that satellites grew more poorly than residents (Table 3). Rather than reflecting conditional development, the small difference in mean body size may be a correlated adaptation to the demands of each reproductive strategy. Independent males fight to establish courts and remain on them throughout the day¹⁸, both of which may favour larger size through increased social dominance and energy storage capacity. In contrast, satellites do not fight for courts, are freer to forage throughout the day, and spend more time flying, reducing the need for, and raising the physiological cost of, maintaining larger body size.

Theoretical and empirical investigations of behavioural development, social behaviour and life history of ruffs are henceforth justified in using a genetic equilibrium model for generating predictions against which to test field observations^{15,20-22}. Our findings highlight the question of what selective conditions favour genetic polymorphism controlling morph development in this species¹⁴, rather than quantitative inheritance²³ or nearly complete environmental control^{24,25}, as occurs in nearly all other species with alternative male mating tactics^{1,3,9,12,13}. □

Received 5 June; accepted 16 August 1995.

1. Dunbar, R. I. M. in *Perspectives in Ethology* Vol. 5 (eds Bateson, P. P. G. & Klopfer, P. H.) 385-431 (Plenum, New York, 1982).

2. Alcock, J. in *Sexual Selection and Reproductive Competition in Insects* (eds Blum, M. S. & Blum, N. A.) 381-402 (Academic, New York, 1979).
3. Arak, A. in *Producers and Scroungers* (ed. Barnard, C. J.) 154-194 (Croom Helm, London, 1984).
4. Zimmerer, E. J. & Kallman, K. D. *Evolution* **43**, 1298-1307 (1989).
5. Ryan, M. J., Pease, C. M. & Morris, M. R. *Am. Nat.* **139**, 21-31 (1992).
6. Erbelding-Denk, C. et al. *Behav. Genet.* **24**, 95-101 (1994).
7. Hogan-Warburg, A. J. *Ardea* **54**, 109-229 (1966).
8. van Rhijn, J. G. *Behaviour* **47**, 153-229 (1973).
9. Caro, T. M. & Bateson, P. *Anim. Behav.* **34**, 1483-1499 (1986).
10. Höglund, J. & Lundberg, A. *Auk* **106**, 336-338 (1989).
11. van Rhijn, J. G. *The Ruff* (Poyser, London, 1991).
12. Dawkins, R. in *Sociobiology: Beyond Nature/Nature?* (eds Barlow, G. W. & Silverberg, J.) 331-367 (Westview, Boulder, CO, 1980).
13. West-Eberhardt, M. J. *A. Rev. Ecol. Syst.* **20**, 249-278 (1989).
14. Lively, C. M. *Am. Nat.* **128**, 561-572 (1986).
15. van Rhijn, J. G. *Ibis* **125**, 482-498 (1983).
16. Shepard, J. M. *Living Bird* **14**, 87-111 (1985).
17. Hill, W. L. *Behav. Ecol. Sociobiol.* **29**, 367-372 (1991).
18. Lank, D. B. & Smith, C. M. *Behav. Ecol. Sociobiol.* **29**, 137-145 (1987).
19. Höglund, J., Montgomerie, R. & Widemo, F. *Behav. Ecol. Sociobiol.* **32**, 31-39 (1993).
20. Slatkin, M. *Am. Nat.* **114**, 384-398 (1989).
21. Austad, S. N. *Am. Zool.* **24**, 309-320 (1984).
22. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, 1982).
23. Cade, W. *Science* **212**, 563-564 (1981).
24. Emlen, D. L. *Proc. R. Soc. Lond. B* **256**, 131-136 (1994).
25. Moran, N. A. *Am. Nat.* **139**, 971-989 (1992).
26. Ricklefs, R. E. in *Avian Biology* Vol. 7 (eds Farner, D. S., King, J. R. & Parkes, K. C.) 1-83 (Academic, New York, 1983).
27. Hanotte, O., Burke, T., Armour, J. A. L. & Jeffreys, A. J. *Genomics* **9**, 587-597 (1991).
28. Bruford, M. W., Hanotte, O., Brookfield, J. F. Y. & Burke, T. in *Molecular Genetic Analysis of Populations: A Practical Approach* (ed. Hoelzel, A. R.) 225-269 (IRL, Oxford, 1992).
29. Dixon, A., Ross, D., O'Malley, S. C. L. & Burke, T. *Nature* **371**, 698-700 (1994).
30. Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. J. *Ann. hum. Genet.* **51**, 269-288.

ACKNOWLEDGEMENTS. We thank A. and R. Ohtonen, E. Pulliainen, other colleagues at the University of Oulu, and Finnair's Technical Services Department for assistance in obtaining our data, and the landowners of Limingänlahti for permission to work with birds nesting along its shores. A. Krupa provided technical assistance with paternity analyses. The research was supported by grants from the US NSF, the H. F. Guggenheim Foundation, the National Geographic Society, the United States Educational Foundation in Finland (Fulbright program), and Queen's University at Kingston (to D.B.L.); the NSERC of Canada (to D.B.L. and F.C.); and the NERC, SERC and the Royal Society (to T.B.).

Cell fate in the *Arabidopsis* root meristem determined by directional signalling

Claudia van den Berg*, Viola Willemsen*, Willem Hage†, Peter Weisbeek* & Ben Scheres*

* Department of Molecular Cell Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

† Netherlands Institute for Developmental Biology, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

POSTEMBRYONIC development in plants is achieved by apical meristems. Surgical studies and clonal analysis have revealed indirectly that cells in shoot meristems have no predictable destiny¹⁻³ and that position is likely to play a role in the acquisition of cell identity⁴⁻⁷. In contrast to animal systems⁸⁻¹⁰, there has been no direct evidence for inductive signalling in plants until now. Here we present evidence for such signalling using laser ablation of cells in the root meristem of *Arabidopsis thaliana*. Although these cells show rigid clonal relationships¹¹, we now demonstrate that it is positional control that is most important in the determination of cell fate. Positional signals can be perpetuated from more mature to initial cells to guide the pattern of meristem cell differentiation. This offers an alternative to the general opinion that meristems are the source of patterning information¹².

The *Arabidopsis thaliana* root consists of single layers of epidermis, cortex, endodermis and pericycle, surrounding a vascular bundle¹³. The root meristem is derived from basal cells of the embryo proper that forms the initials for the different cell types present in the mature root, and from the hypophysis which generates the columella root cap and quiescent centre¹¹ (Fig. 1). The consistency in cell fate and position suggests that cells are

committed to form particular tissue types. This is supported by two β -glucuronidase promoter fusions specific for vascular and root cap cells, which are also expressed in the meristematic initial cells (Fig. 2a, b).

To determine whether root meristem cells behave according to their clonal origin or to their position, we applied laser ablation. Dead cells are compressed towards the periphery of the root (Fig. 3a), allowing neighbouring cells to invade their position. By recording the fate of the invading cell, the contribution of lineage or position can be investigated.

Upon ablation of all quiescent centre cells, the dead cells become flattened and are displaced towards the root tip. Cells of the proximal vascular bundle occupy the former position of the dead cells. These cells no longer express the vascular marker (Fig. 2c) but instead express the root cap marker (Fig. 2d). Hence the clonal boundary set by the first zygotic division, separating future vascular and root cap cells, does not restrict developmental potential. Furthermore, information guiding cell fate along the apical-basal axis in the root tip is permanently present. These results confirm earlier regeneration studies^{14,17} but also show that global redifferentiation is not necessary for cell-fate switching.

We investigated whether positional information can determine cell fate in the radial plane. Cortical initials divide asymmetrically to produce cortex/endodermis cells (Fig. 3a). Upon ablation of these initials, pericycle cells invade and divide periclinally. This maintains the pericycle cell file and generates new cells in the cortical cell file (Figs 3a, 4b). Pericycle cells are smaller than cortical cells, so more than one pericycle cell invades, thus enlarging the typical number of eight cortical cells. The former pericycle cells subsequently divide asymmetrically (Figs 3b, 4b), generating cortical and endodermal cell files. Suberin staining revealed a casparian strip in all prospective endodermal cells, illustrating their differentiation to endodermal fate (Fig. 3c). Therefore, pericycle cells switch fate when moved over radial clonal boundaries.

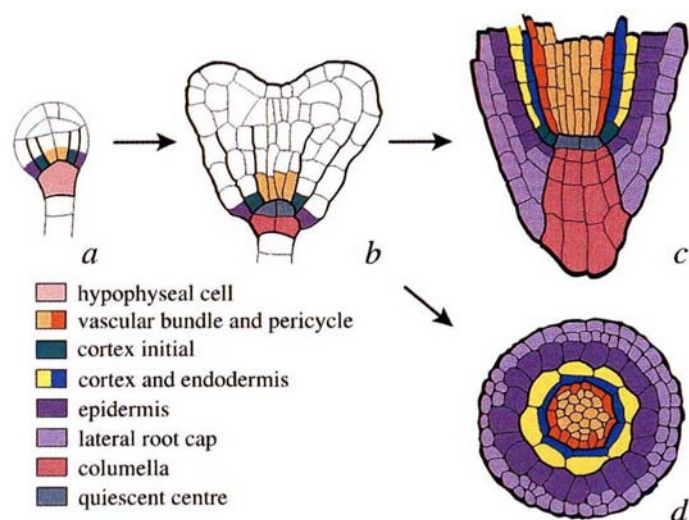


FIG. 1 Schematic representation of the *Arabidopsis* root meristem origin^{11,13}. a, Early globular embryo; b, early heart-stage embryo; c, seedling root (longitudinal); d, seedling root (transverse). The root meristem is derived from the basal cells of the embryo proper and from the hypophysis (a). The initial cells for the different tissue types are separated early during embryogenesis. The cortical initial cells first divide anticlinally, generating cortical daughters, whereupon these divide periclinally to give rise to cortex and endodermis. Epidermal initials form epidermis and lateral root cap. The hypophysis gives rise to the columella and the quiescent centre. The four quiescent centre cells cease dividing after formation¹³.

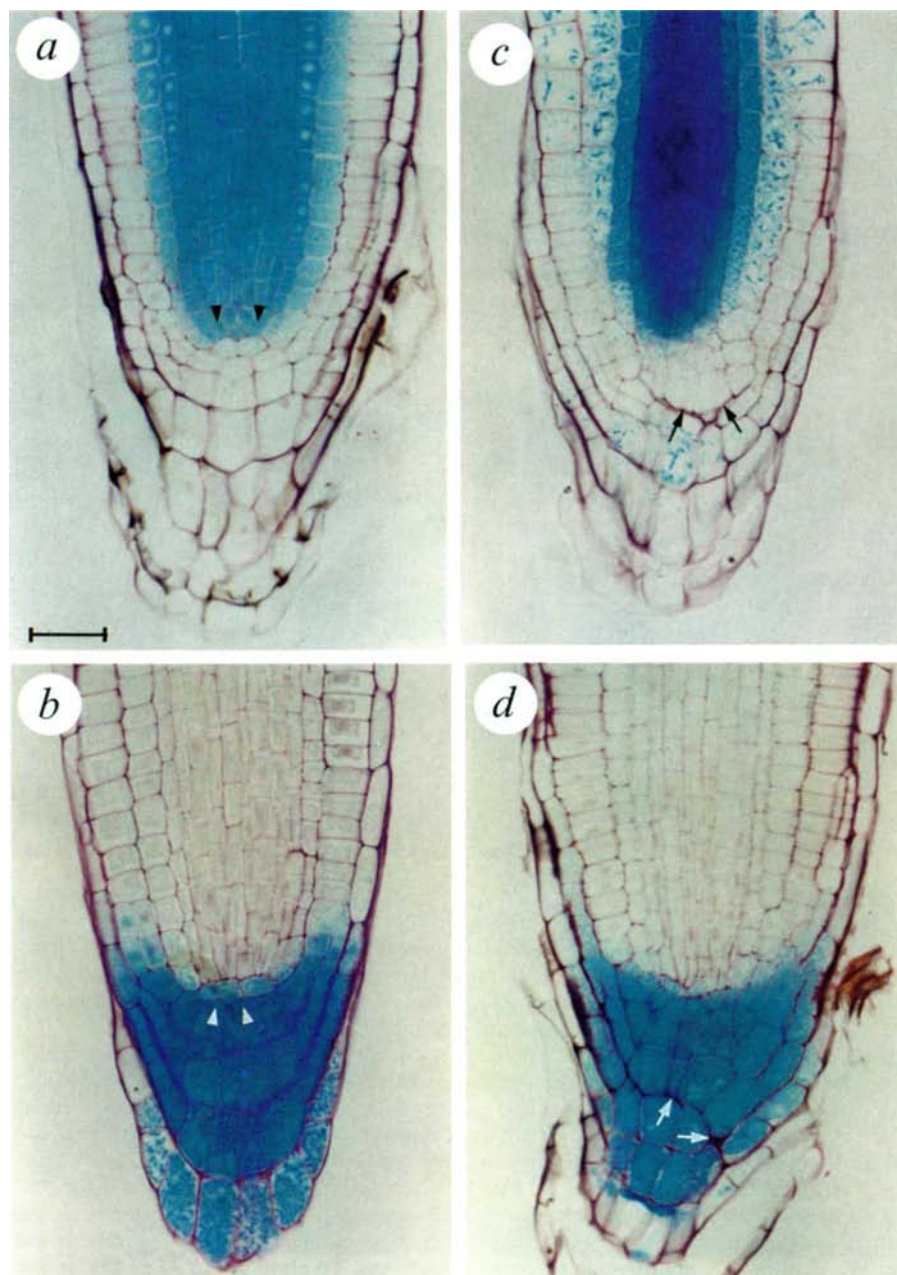
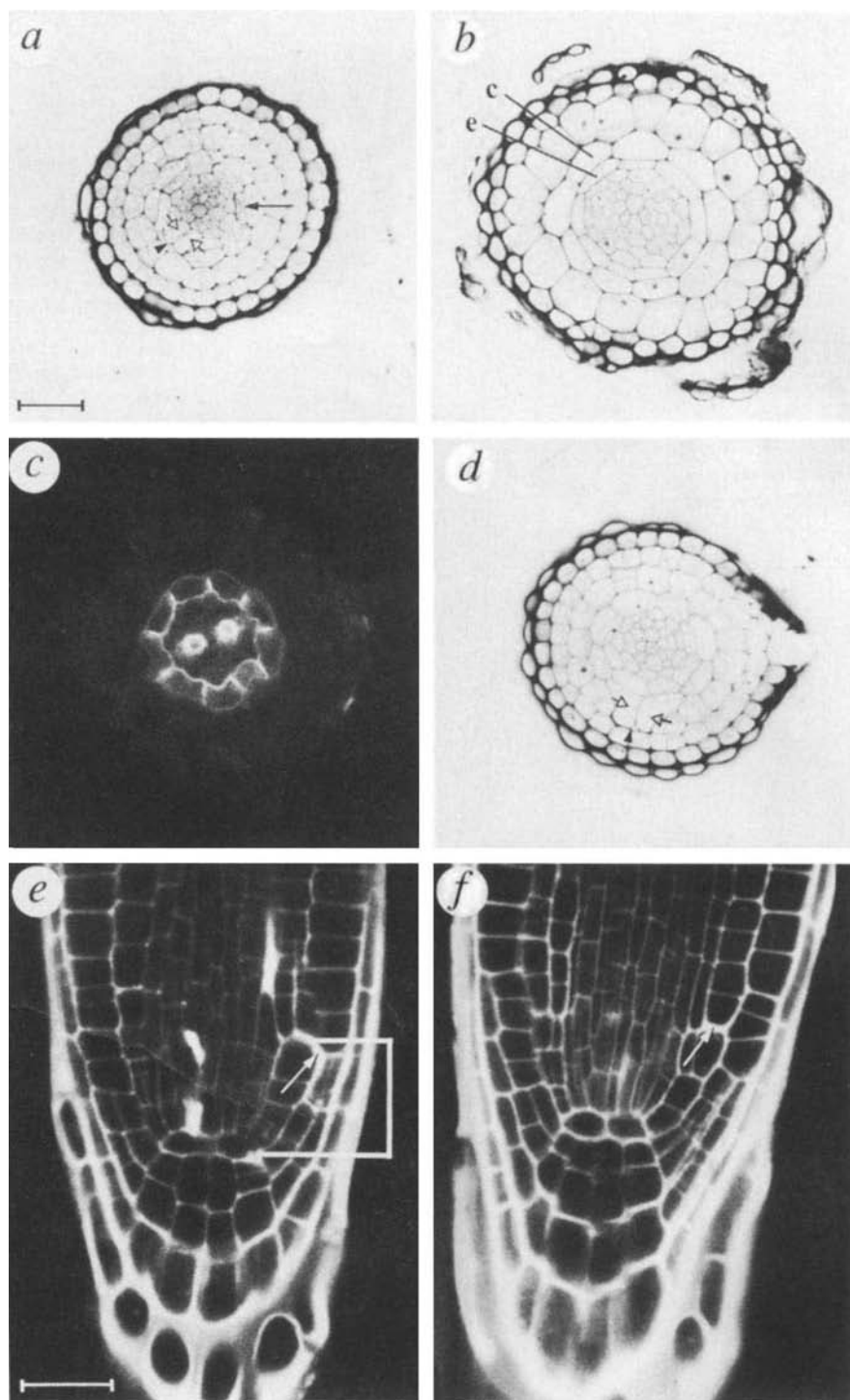


FIG. 2 Allocation of fate is determined by cell position rather than by clonal origin. β -Glucuronidase reporter gene expression in a, root vascular tissue, and b, root cap. Note the expression in initial cells (arrowheads). c, d, Expression of reporter genes 5 days after ablation of the quiescent centre. Dead cells are displaced towards the root tip (arrows; $n > 50$). Cells at the former position of the quiescent centre form continuous files with the vascular bundle ($n > 50$). These cells always cease to express the vascular marker (c) ($n = 8$). In contrast, they now express the root-cap marker (d) ($n = 13$). Thus cells that cross the clonal boundary switch fate according to positional cues. Scale bar, 25 μ m.

METHODS. Two-day-old seedlings were incubated with 10 μ g ml⁻¹ propidium iodide (Sigma) and examined using a confocal laserscan microscope (25 mW argon-ion laser, mrc-600, Bio-Rad, Zeiss Axiovert). Propidium iodide surrounds living cells, thereby outlining individual cells (compare with Fig. 3e) and presensitizes the cells for ablations (not shown). Individual cells were ablated by focusing the unfiltered laser beam for 1 s on a specific position and plane of focus. Successfully ablated cells were stained by the entry of propidium iodide. The ablated cells could be retained following growth. Remnants of ablated cells retained a fluorescent signal after fixation and embedding (not shown). Successfully ablated roots, grown for various periods, were stained in 0.5 mg ml⁻¹ X-gluc (Biosynth AG) for 2 h, fixed overnight in 70% ethanol. Dehydration, embedding and sectioning were performed as described¹¹. The vascular-marker line (553–643) contains an enhancer trap construct²⁶. The root-cap marker line contains the 35S CAMV promoter B2 subdomain fused to the β -glucuronidase gene²⁷.

FIG. 3 Positional information is generated from cells within the same tissue type and is perpetuated distally. *a*, Ablation of one cortical initial cell. The dead cell is pressed peripherally (arrowhead) and two pericycle cells are occupying its former position ($n=35$). These have divided periclinally to generate cells in both the cortical and pericycle layer (open arrows). Note the asymmetric division in untreated cortical initials (arrow). *b*, Three days after ablation of a cortical initial. The invading pericycle cells have generated a larger cortical (*c*) and a smaller endodermal cell (*e*) ($n=15$). Two pericycle cells occupy the position of one dead cortical cell, forming 9 cortical and endodermal cells. *c*, Fresh section of a root at a similar stage as that in *b*. Suberin staining shows the endodermis-specific casparian strip in all nine endodermal cells ($n=7$). *d*, Section of a root, one day after an epidermal initial cell ablation. The dead cell is compressed peripherally (arrowhead). The invading cortical initial has divided to form new cells in the epidermal layer (arrows; $n=21$). These cells later generate new root cap cells ($n=8$; not shown). *e*, Confocal laser scan of a root 4 days after ablation of 3 cortical initial daughters. The cell file (vertical bar) of at least 3 ($n=13$) or even 4 ($n=6$) cells under the dead fluorescent cell (arrow) arises from one initial which has divided anticlinally and not periclinally. *f*, Scan of a root 4 days after ablation of one epidermal, one cortical and one pericycle daughter. The dead cells are compressed (arrow) and the underlying initial has undergone normal anti- and periclinal divisions ($n=11$). Scale bars, 25 μm .

METHODS. Sections were stained with 0.1% Astra blue (Merck) and suberin-stained as described^{11,25}. For ablation of 3 cortical daughters, a c-apochromat 40 $\times$ /1.2 W corr. lens (Zeiss) was used to score successful ablations by making an X-Z-scan. If the dead cells were rapidly compressed to the root periphery, proximal signalling was restored, before the cortical initial could perform a number of divisions ($n=17$). If the dead cells stayed in place, always 3 or 4 cells without asymmetric division were seen under the compressed dead cell ($n=19$).



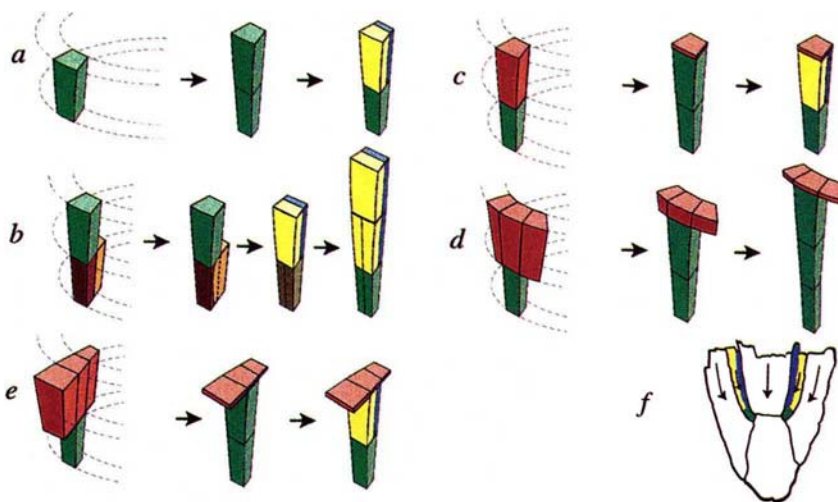
Epidermal initials generate the epidermal cell file but also form the lateral root cap (Fig. 1). After ablation of an epidermis initial, a neighbouring cortical cell occupies its position. It divides and generates a daughter in the epidermal layer that performs a characteristic periclinal division (Fig. 3*d*). Upon prolonged growth, intact root cap and epidermis layers arise. Therefore, besides pericycle, cortical cells also respond to changes in position.

Surgical experiments have revealed flexibility of cell fate in both root and shoot apices, but did not demonstrate the exact mode of regeneration and the cellular origin of flexibility¹⁶. Clonal analysis has demonstrated cell flexibility indirectly¹⁸. Here, we show directly that a shift of cells into a different position is the only causal necessity for changing fate. We demon-

strate that meristem cells change fate, despite the partial commitment indicated by marker-gene expression. All cell types that we tested can adopt position-dependent fates, including the quiescent centre (not shown). It is therefore tempting to regard the root meristem as a single equivalence group^{19,20}.

If initial cells behave according to position, continuous signalling should maintain the functional pattern. We used the cortical initial cell to probe the direction of signals. Daughters of cortical initial cells before their asymmetric periclinal division, were ablated and the behaviour of the underlying cortical initial cell was monitored. Ablation of one cortical daughter cell had no effect on the sequence of divisions occurring in the initial cell (Fig. 4*c*). However, this cortical initial still abuts cortical daughters through three-way junctions. To eliminate putative proximal

FIG. 4 Schematic representations of cortical initial and derivative cell ablation experiments. Ablated cells are depicted in red; the colours of the cells are as in Fig. 1. *a*, Position of cortical initial cell and derivatives in one of eight identical meristem cell files. The cortical initial divides transversally, generating a cortical daughter cell which subsequently divided periclinally to form an endodermal and cortical cell file (97%; $n=215$). *b*, Ablation of a cortical initial. The dead cell is compressed to the outside and its position is filled with pericycle cells. These switch fate and behave as cortical initials. *c*, Ablation of one cortical initial daughter cell. The underlying cortical initials perform normal periclinal and anticlinal cell divisions ($n=16$). Note that the cortical initial is not isolated from all cortex daughter cells and passage of signals from the daughters of neighbouring initials to the middle initial can still occur. *d*, The isolation of one cortical initial upon ablation of 3 cortical daughter cells. The cortical initial is still able to perform proliferative divisions to generate daughter cells, but asymmetric divisions are absent. *e*, Control experiment in which one epidermal, one cortical daughter and one pericycle cell above one cortical initial are ablated. A similar number of dead cell remnants contacting one initial cell are present (as in *d*), but the initial is not isolated from all the cortical daughters. The underlying initial performs its normal division pattern.



f, Model representing pattern perpetuation by positional signals in the root meristem. Colours represent cortical/endodermal cell layer for which we have demonstrated directional flow of positional cues (arrows). Other arrows indicate that signals can similarly be perpetuated within other individual cell layers.

signals in the cortical cell file, we ablated three cortical daughter cells and scored the effects on the underlying, now isolated, initial cell. This initial cell could still generate a daughter cell but was unable to divide asymmetrically to generate the cortex and endodermis files, generating a single file of cells instead (Figs 3*e*, 4*d*). To eliminate the possibility that this is a wounding effect, we ablated three neighbouring cells, but this time one cortical, epidermal and pericycle daughter cell. In this case, one cortical initial cell abutted three dead cells but was not isolated from other cortical cells. Here, the cortical initial cell performed its normal anti- and periclinal divisions (Figs 3*f*, 4*e*). We conclude that information guiding allocation of cell fate in the radial plane is propagated through an individual cell layer and is directed towards the tip (Fig. 4*f*). Thus in the root meristem, inductive processes appear to specify initial cells, whereby the mature daughter cells act as a patterning template. Similarly, in *Xenopus laevis* when single cells are transplanted to heterotopic sites, they differentiate to conform with their neighbouring host cells²¹. Our results indicate that meristem initial cells are not generators of pattern, as is generally assumed¹², but rather perpetuate an existing pattern.

Arabidopsis mutants that lack asymmetric cortical divisions offer a potential starting point for unravelling the molecular mechanisms of position-dependent inductive signalling in plants^{22,23}. Experiments on *Fucus* embryos have indicated that cell wall components are involved in cell fate decisions²⁴. Alternatively, in the *Arabidopsis* root at least one cell layer is symbolically coupled and so cytoplasmic inductive signals might be restricted to cell layers²⁵. □

17. Schiavone, F. M. & Racusen, R. H. *Development* **113**, 1305–1313 (1991).
18. Poethig, R. S. *Am. J. Bot.* **74**, 581–594 (1987).
19. Sulston, J. E. & White, J. G. *Dev. Biol.* **78**, 577–597 (1980).
20. Kimble, J. *Dev. Biol.* **87**, 286–300 (1981).
21. Haesman, J., Wylie, C. C., Hausen, P. & Smith, J. C. *Cell* **37**, 185–194 (1984).
22. Benfey, P. N. *et al. Development* **119**, 57–70 (1993).
23. Scheres, B. *et al. Development* **121**, 53–62 (1995).
24. Berger, F., Taylor, A. & Brownlee, C. *Science* **263**, 1421–1423 (1994).
25. Duckett, C. M., Oparka, K. J., Prior, D. A. M., Dolan, L. & Roberts, K. *Development* **120**, 3247–3255 (1994).
26. Goddijn, O. J. M., Lindsey, K., van der Lee, F. M., Klap, J. C. & Sijmons, P. C. *Plant J.* **4**, 863–873 (1993).
27. Benfey, P. N., Ren, L. & Chua, N.-H. *EMBO J.* **9**, 1685–1696 (1990).

ACKNOWLEDGEMENTS. We thank P. Sijmons and MOGEN for the vascular-specific marker line 553–643; P. Benfey for the root-cap marker; F. Kindt, R. Leitho, W. Veenendaal and P. Brouwer for photography; D. Smit and F. Gerards for artwork; I. Sussex for his suggestions and H. McKhann for critically reading the manuscript.

Salt-resistant hypertension in mice lacking the guanylyl cyclase-A receptor for atrial natriuretic peptide

M. James Lopez^{†‡}, Stephen K.-F. Wong[†], Ichiro Kishimoto^{†‡}, Susan Dubois^{†§}, Valerie Mach^{†‡}, Janet Friesen^{†‡}, David L. Garbers^{†‡} & Annie Beuve^{†‡}

[†] The Howard Hughes Medical Institute and the Departments of [‡] Pharmacology, ^{*} Pediatrics and [§] Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9050, USA

AROUND half of all humans with essential hypertension are resistant to salt (blood pressure does not change by more than 5 mm Hg when salt intake is high)^{1–5}, and although various inbred strains of rats display salt-insensitive elevated blood pressure⁶, a gene defect to account for the phenotype has not been described. Atrial natriuretic peptide (ANP) is released from the heart in response to atrial stretch and is thought to mediate its natriuretic and vaso-relaxant effects through the guanylyl cyclase-A receptor (GC-A)⁷. Here we report that disruption of the GC-A gene results in chronic elevations of blood pressure in mice on a normal salt diet. Unexpectedly, the blood pressure remains elevated and unchanged

Received 25 July; accepted 29 August 1995.

1. Pilkington, M. *New Phytol.* **28**, 37–53 (1929).
2. McDaniel, C. N. & Poethig, R. S. *Planta* **175**, 13–22 (1988).
3. Jegla, D. E. & Sussex, I. M. *Dev. Biol.* **131**, 215–225 (1989).
4. Wardlaw, C. W. *Phil. Trans. Soc. London B* **234**, 583–604 (1950).
5. Stewart, R. N. & Dermen, H. *Am. J. Bot.* **57**, 816–826 (1975).
6. Steeves, T. A. & Sussex, I. M. in *Patterns of Plant Development* 2nd edn 86–99 (Cambridge Univ. Press, New York, 1989).
7. Dawe, R. K. & Freeling, M. *Plant J.* **1**, 3–8 (1991).
8. Spemann, H. & Mangold, H. *Roux's Arch. dev. Biol.* **100**, 599–638 (1924).
9. Greenwald, I. & Rubin, G. M. *Cell* **68**, 271–281 (1992).
10. Jessell, T. M. & Melton, D. A. *Cell* **68**, 257–270 (1992).
11. Scheres, B. *et al. Development* **120**, 2475–2487 (1994).
12. Alberts, B. *et al. in Molecular Biology of the Cell* 3rd edn 1108–1113 (Garland, New York, 1994).
13. Dolan, L. *et al. Development* **119**, 71–84 (1993).
14. Barlow, P. *New Phyt.* **73**, 937–954 (1974).
15. Feldman, L. J. & Torrey, J. G. *Am. J. Bot.* **63**, 345–355 (1976).
16. Sussex, I. M. *Cell* **56**, 225–229 (1989).

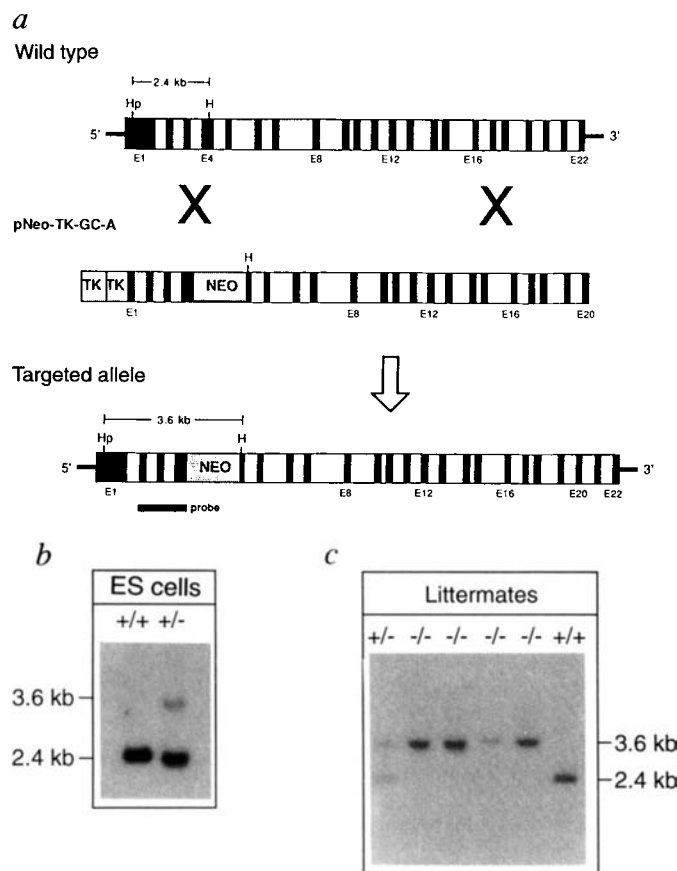


FIG. 1 Targeted disruption of the mouse GC-A gene. **a**, Map of the wild-type GC-A allele (top), the targeting vector pNeoTK-GC-A (middle) and predicted disrupted allele (bottom). Black boxes indicate exons. The targeting vector was made by placing a *neo*^r cassette (Neo) into exon 4 that corresponds to a part of the extracellular putative ligand-binding domain. TK, thymidine kinase. Horizontal lines above the wild-type allele and the predicted disrupted allele represent the size of expected bands of an *HpaI*-*HindIII* restriction digest. The small bar represents the probe used for detection on Southern blots, H, *HindIII*; Hp, *HpaI*. **b**, Southern blot analysis of control (AB1) and targeted (4A11) ES cell clones with an *HpaI*-*HindIII* digest. The position of the bands for the corresponding wild-type (2.4 kilobases) and mutant (3.6 kb) alleles are indicated. **c**, Southern blot analysis of littermates from a heterozygous intercross of F₁ mice. Hybridization with a probe extending over exons 2-4 showed the expected 2.4-kb band for wild-type mice, 2.4-kb and 3.6-kb bands for heterozygotes and only the expected 3.6-kb band for homozygous mutant mice.

METHODS. A genomic library of the 129/SVJ mouse strain was used to amplify genomic DNA fragments by the polymerase chain reaction (PCR). These fragments covered a total of 10 kb of the mouse genomic sequence from exon 1 to exon 20 (based on rat GC-A gene structure)²³. The fragments were inserted into the plasmid pPollshort-neobPA-HSVTK (ref. 24) in which the unique *Bam*HI site was changed to a *Not*I site. The targeting vector (pNeo-TK-GC-A) contained an insertion of the neomycin-resistance gene into exon 4 of the GC-A gene and two tandem copies of HSVTK (denoted TK in **a**) at the 5' end of the mouse GC-A sequence. ES cells were electroporated at 230 V, 500 μ F with the linearized plasmid and placed under selection with G418 (0.18 mg ml⁻¹) and FIAU (0.2 μ M) on an irradiated STO-LIF fibroblast feeder layer²⁵. The ES cells were expanded and genomic DNA prepared for PCR analysis²⁶. The clones were first screened by PCR with a primer outside the targeting construct in exon 1 (5'-TGTCAGCTGGGCGACTTCGTGACG-3') and a second primer in the *neo*^r cassette (5'-GATTGGGAAGACAATAG-CAGGCATGC-3') which should amplify a 2.4-kb PCR fragment in successfully targeted cells. The positive clones were then expanded and screened by Southern blot analysis with an *HpaI*-*HindIII* digest. Two correctly targeted clones were injected into C57BL/6J blastocysts, which were implanted in pseudopregnant females to generate chimaeric males. Four of these chimaeras transmitted the disrupted GC-A gene through the germ line. F₁ mice (+/-) having the GC-A disruption were bred to obtain +/+, +/-, -/- F₂ offspring.

in response to either minimal or high salt diets. Aldosterone and ANP concentrations are not affected by the genotype. Therefore, mutations in the GC-A gene could explain some salt-resistant forms of essential hypertension and, coupled with previous work⁸, further suggest that the GC-A signalling pathway dominates at the level of peripheral resistance, where it can operate independently of ANP.

A targeting construct was designed to disrupt the GC-A gene by inserting the neomycin-resistance (*neo*^r) gene into exon 4 (Fig. 1a). GC-A contains a single transmembrane domain that divides a ligand-binding domain from intracellular protein kinase-like and cyclase catalytic domains⁹, and exon 4 lies within the amino-terminal ligand-binding region. The construct was transfected into murine G1 embryonic stem (ES) cells derived from the 129/SVJ mouse strain. Two independent clones containing the predicted mutant allele were selected to produce chimaeric male mice (Fig. 1b), four of which transmitted the disrupted gene through the germ line in matings with female C57BL/6J mice.

The heterozygotes were intercrossed and the genotype of the F₂ offspring was analysed by Southern blot at 3 weeks (Fig. 1c). The genotype ratio of +/+ : +/- : -/- was 1/1.83/0.70, based on a total of 283 F₂ and F₃ offspring analysed. The slight reduction in the number of homozygous mutant mice expected is statistically significant ($P < 0.001$, χ^2) and is accounted for, at least in part, by approximately 10% fetal hydrops among homozygous embryos (data not shown).

There was no stable transcript of the size of GC-A in the homozygous null, and about 50% of the amount of messenger RNA of wild-type mice in the heterozygous liver and kidney (data not shown). Also, the hearts, kidneys and vasculature of the homozygous mutants (aged less than 5 months) were normal when examined by histological methods.

It has been suggested that ANP released from the heart in response to atrial stretch binds to a guanylyl cyclase-coupled receptor (GC-A) in the kidney to mediate natriuresis and diuresis, and to the same receptor in the vasculature to mediate relaxation⁷. Both homozygous and heterozygous GC-A-mutant mice displayed an elevated systolic blood pressure (27.4 mm Hg and 10.5 mm Hg, respectively) compared with the wild-type controls when fed a diet containing 0.7% NaCl (Fig. 2). Similar increases also occurred in mean blood pressure (+/-, 7.7 mm Hg; -/-, 19.7 mm Hg) and diastolic blood pressure (+/-, 5.9 mm Hg; -/-, 15.3 mm Hg) (Fig. 2). Heart rates did not differ significantly between the groups (mean beats per min $\pm$ s.e.m.; +/+, 674 $\pm$ 12; +/-, 670 $\pm$ 11; -/-, 659 $\pm$ 14).

The natriuretic peptide/cyclase signalling pathway has often been thought of as an acute sensor of changes in blood volume, opposing the actions of the renin/angiotensin/aldosterone axis^{10,11}. The previous work on disruption of the ANP gene supported this model, because ANP-depleted mice displayed normal blood pressure unless stressed with high salt⁸. Disruption of the GC-A gene was therefore expected to lead to salt-sensitive forms of hypertension. However, blood pressure and heart rates remained unchanged in mice whose diets contained either 0.008% or 8% NaCl (Fig. 3), suggesting that the mice were able to respond to the changes in salt intake and maintain relative salt homeostasis.

The adrenal gland seemed to respond appropriately to the level of NaCl intake in either the mutant or wild-type mice, as plasma aldosterone was elevated (mean $\pm$ s.d.: +/+, 3,700 $\pm$ 1,580 pmol l⁻¹, n = 12; +/-, 3,910 $\pm$ 960 pmol l⁻¹, n = 11; -/-, 5,580 $\pm$ 1,800 pmol l⁻¹, n = 11) on the 0.008% diet. The high-salt diet reduced aldosterone to less than 140 pmol l⁻¹ in all three genotypes. There were also no effects of genotype on serum ANP concentrations, haematocrit or body weight (data not shown). The failure of the high-salt diet to change haematocrit or body weight, indicators of changes in blood volume, suggests that the kidney maintained the constant blood pressure through increased excretion of salt in the urine.

The disruption of the GC-A gene leads to new concepts with respect to natriuretic peptide/guanylyl cyclase regulation of blood pressure. GC-A seems to exert its major effects at the level of the vasculature, and does so independently of salt. Disruption of the gene for the putative ligand, ANP, failed to yield an elevated blood pressure in animals on a normal diet in previous studies⁸. One explanation for the difference in phenotype is that GC-A binds ligands other than ANP. In fact, a natriuretic peptide known as B-type natriuretic peptide (BNP)¹² binds to GC-A with an affinity comparable to that of ANP¹³. BNP is released principally from the ventricles of the heart¹⁴, possibly in a constitutive fashion, and could therefore represent a chronic regulator of blood pressure. Another explanation for a difference in phenotype between the ANP and GC-A gene knockouts is that basal cyclase activity of GC-A is sufficient to maintain normal vascular resistance in the absence of any ligand.

The finding that the GC-A gene knockout does not yield a salt-sensitive form of hypertension could be explained if ANP regulates natriuresis through a receptor other than GC-A. ANP, in fact, is known to bind to a truncated member of the guanylyl cyclase family known as the ANP clearance receptor¹⁵. The clearance receptor has been suggested to clear ANP from the circulation by some¹⁶, and to act as a signalling molecule by others^{17,18}. ANP also binds with low affinity to a second guanylyl cyclase-coupled receptor known as GC-B¹⁹.

The further finding that peripheral resistance is affected by the disruption of the GC-A gene without a chronic effect on kidney function agrees in principal with previous work on transgenic mice in which high expression of ANP as a transgene led to decreased blood pressure in the face of apparently normal kidney function (natriuresis and diuresis)²⁰. Volume expansion

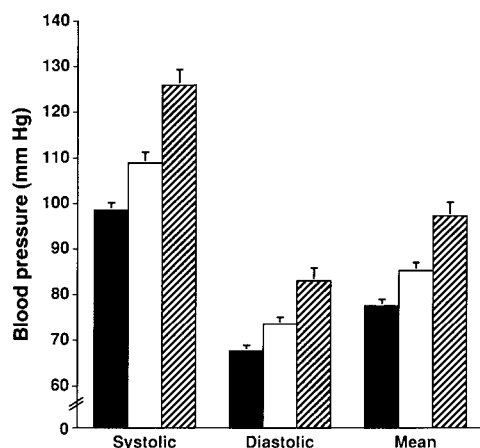


FIG. 2 The effect of disruption of the GC-A gene on systolic, diastolic and mean blood pressure (BP). All experiments were carried out in a blinded fashion on mice eating a standard rodent chow (0.7% NaCl, Harlan-Teklad; wild-type, $n=12$ (black bars), heterozygous mice, $n=12$ (open bars) and homozygous mutant mice, $n=12$ (striped bars)). Results are shown as mean $\pm$ s.e.m. and statistics were by ANOVA (P -values) and Tukey's test of the means (P -values). Tail cuff pressures of F_2 mice aged 90–180 days were measured with a computerized, automated system (Softron, Tokyo, Japan) similar to systems previously described^{8,27,28}. Recent work has shown a strong correlation between tail cuff estimates and direct intra-arterial measurements of blood pressure²⁷. Mice were age-matched between the groups and were trained for 7 days before experimental measurements. Three to six measurements per day for four consecutive days were taken for each mouse. Significant differences were observed between the genotypes ($P<0.001$; systolic BP: $-/-$ compared with $+/+$, $P<0.001$; $-/-$ compared with $+/-$, $P<0.005$; mean BP: $-/-$ compared with $+/+$, $P<0.001$; $-/-$ compared with $+/-$, $P<0.025$; diastolic BP: $-/-$ compared with $+/+$, $P<0.001$; $-/-$ compared with $+/-$, $P<0.005$; differences between $+/-$ and $+/+$ were not significant), but not between males and females of the same genotype (data not shown).

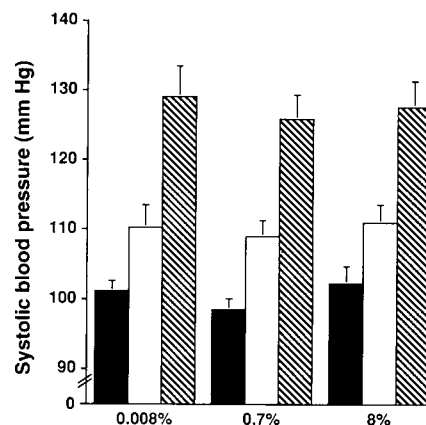


FIG. 3 Systolic blood pressure in mice on high, normal and low salt diets. Wild-type (solid bars), heterozygous (open bars) and homozygous mutant mice (striped bars) fed one of three diets from Harlan-Teklad for 2 weeks are shown: 0.008% NaCl, ($+/+$, $n=12$; $+/-$, $n=13$; $-/-$, $n=11$), 0.7% NaCl ($+/+$, $n=12$; $+/-$, $n=12$; $-/-$, $n=12$) and 8% NaCl ($+/+$, $n=6$; $+/-$, $n=6$; $-/-$, $n=6$). Blood pressure was measured on alternating days throughout the treatment period and on 2–3 consecutive days at the end of the 2-week period. Six to twelve estimates of blood pressure on each individual were made on days of measurement. On days when blood pressure was not measured the animals underwent training. The values represent the means $\pm$ s.e.m. Statistical analysis was by ANOVA (P -values) and Tukey's test of the means (P -values). There were significant differences in blood pressure between the genotypes on all diets ($P<0.001$; 0.008% NaCl: $-/-$ compared with $+/+$, $P<0.001$; $-/-$ with $+/-$, $P<0.005$; 8% NaCl: $-/-$ with $+/+$, $P<0.001$; $-/-$ with $+/-$, $P<0.005$; there were no significant differences between $+/-$ and $+/+$ mice), but no significant differences between the three salt diets. Blood pressure was measured in the mice on the 8% salt diet once again at 3 and 4 weeks and no change was observed (data not shown).

continued to cause a natriuretic and diuretic effect in the transgenic animals.

A unifying explanation for the mice expressing high levels of ANP as a transgene²⁰, the ANP gene disruption⁸ and our studies is that the heart releases a natriuretic activity other than ANP/BNP, and that a receptor other than GC-A is the target receptor for this activity in the kidney. The 'other' natriuretic activity could reside in another part of the ANP prohormone or be a completely different entity. Disruption of the ANP gene also eliminated the granules in the heart⁸, and thus other natriuretic-like factors found in these same granules would have been eliminated. The possibility that natriuretic activity of the prohormone of ANP explains the differences in phenotype in the various studies is supported by reports showing that ANP causes decreases in blood pressure, increases in sodium excretion and elevations of cyclic GMP, whereas pro-ANP (amino acids 31–67) has no effect on blood pressure or cyclic GMP, but induces natriuresis²¹. As the ANP gene disruption would have also eliminated pro-ANP, salt sensitivity would be evident in this model if pro-ANP regulates natriuresis through a receptor other than GC-A.

The phenotype of the GC-A gene knockout resembles that of almost 50% of human patients with essential hypertension in being salt-resistant. The apparently normal regulation of aldosterone is also seen in humans with salt-resistant essential hypertension²². Therefore, GC-A gene defects may account for, or exacerbate, some salt-insensitive forms of hypertension. $\square$

Received 26 June; accepted 11 September 1995.

1. Oparil, S. et al. *J. cardiovasc. Pharmacol.* **12**, S56–S69 (1988).
2. Weinberger, M. H. in *Hypertension: Pathophysiology, Diagnosis, and Management* (eds Laragh, J. H. & Brenner, B. M.) 1999–2010 (Raven, New York, 1990).
3. Weinberger, M. H., Miller, J. Z., Lust, F. C., Grim, C. E. & Fineberg, N. S. *Hypertension* **8** (Suppl II), II-127–II-134 (1986).

4. Kawasaki, T., Deka, C. S., Bartter, F. C. & Smith, H. *Am. J. Med.* **64**, 193–198 (1978).
5. Fujita, T., Henry, W. L., Bartter, F. C., Lake, C. R. & Delea, C. S. *Am. J. Med.* **69**, 334–344 (1980).
6. Kimura, G., Frem, G. J. & Brenner, B. M. *Curr. Nephrol. Hypertension* **3**, 1–12 (1994).
7. Drewett, J. & Garbers, D. L. *Endocrine Rev.* **15**, 135–162 (1994).
8. John, S. W. M. et al. *Science* **267**, 679–681 (1995).
9. Chinkers, M. et al. *Nature* **338**, 78–83 (1989).
10. Atlas, S. A. & Laragh, J. H. in *Hypertension: Pathophysiology, Diagnosis, and Management* (eds Laragh, J. H. & Brenner, B. M.) 861–883 (Raven, New York, 1990).
11. Ballermann, B. J., Zeidel, M. L., Gunning, M. E. & Brenner, B. M. in *The Kidney* (eds Brenner, B. M. & Rector, F. C.) 510–583 (Saunders, Philadelphia, 1991).
12. Sudoh, T., Kangawa, K., Minamino, N. & Matsuo, H. *Nature* **332**, 78–81 (1988).
13. Suga, S.-I. et al. *Endocrinology* **130**, 229–239 (1992).
14. Ogawa, Y. et al. *J. clin. Invest.* **93**, 1911–1921 (1994).
15. Fuller, F. et al. *J. biol. Chem.* **263**, 9395–9401 (1988).
16. Nussenzveig, D. R., Lewicki, J. A. & Maack, T. *J. biol. Chem.* **265**, 20952–20958 (1990).
17. Levin, E. R. *Am. J. Physiol.* **264**, E483–E489 (1993).
18. Anand-Srivastava, M. B. & Trachte, G. J. *Pharmac. Rev.* **45**, 455–497 (1993).
19. Koiler, K. J. et al. *Science* **252**, 120–123 (1991).
20. Steinhilper, M. E., Cochran, K. L. & Field, L. J. *Hypertension* **16**, 301–307 (1990).
21. Dietz, J. R., Vesely, D. L. & Nazian, S. J. *Clin. exp. Pharmac. Hypertension* **21**, 599–606 (1994).
22. Sullivan, J. M. & Ratts, T. E. *Hypertension* **11**, 717–723 (1988).
23. Yamaguchi, M., Rutledge, L. J. & Garbers, D. L. *J. biol. Chem.* **265**, 20414–20420 (1990).
24. Ishibashi, S. et al. *J. clin. Invest.* **92**, 883–893 (1993).
25. Ramirez-Solis, R., Davis, A. C. & Bradley, A. *Meth. Enzym.* **225**, 855–878 (1993).
26. Soriano, P., Montgomery, C., Geske, R. & Bradley, A. *Cell* **64**, 693–702 (1991).
27. Kregge, J. H., Hodgson, J. B., Hagaman, J. R. & Smithies, O. *Hypertension* **25**, 1111–1115 (1995).
28. Fukamizu, A. et al. *J. biol. Chem.* **268**, 11617–11621 (1993).

ACKNOWLEDGEMENTS. M.J.L., S.K.-F.W. and A.B. are equivalent authors on the manuscript. This work was supported in part by grants from the NIH.

A sex difference in the human brain and its relation to transsexuality

Jiang-Ning Zhou*, Michel A. Hofman*,
Louis J. G. Gooren† & Dick F. Swaab*‡

* Graduate School Neurosciences Amsterdam,
Netherlands Institute for Brain Research, Meibergdreef 33,
1105 AZ Amsterdam Z.O., The Netherlands

† Department of Endocrinology, Free University Hospital,
1007 MB Amsterdam, The Netherlands

TRANSSEXUALS have the strong feeling, often from childhood onwards, of having been born the wrong sex. The possible psychogenic or biological aetiology of transsexuality has been the subject of debate for many years^{1,2}. Here we show that the volume of the central subdivision of the bed nucleus of the stria terminalis (BSTc), a brain area that is essential for sexual behaviour^{3,4}, is larger in men than in women. A female-sized BSTc was found in male-to-female transsexuals. The size of the BSTc was not influenced by sex hormones in adulthood and was independent of sexual orientation. Our study is the first to show a female brain structure in genetically male transsexuals and supports the hypothesis that gender identity develops as a result of an interaction between the developing brain and sex hormones^{5,6}.

Investigation of the genetics, gonads, genitalia or hormone level of transsexuals has not, so far, produced any results that explain their status^{1,2}. In experimental animals, however, the same gonadal hormones that prenatally determine the morphology of the genitalia also influence the morphology and function of the brain in a sexually dimorphic fashion^{6,7}. This led to the hypothesis that sexual differentiation of the brain in transsexuals might not have followed the line of sexual differentiation of the body as a whole. In the past few years, several anatomical differences in relation to sex and sexual orientation have been observed in the human hypothalamus (see ref. 6 for a review), but so far no neuroanatomical investigations have been made in relation to the expression of cross-gender identity (transsexuality).

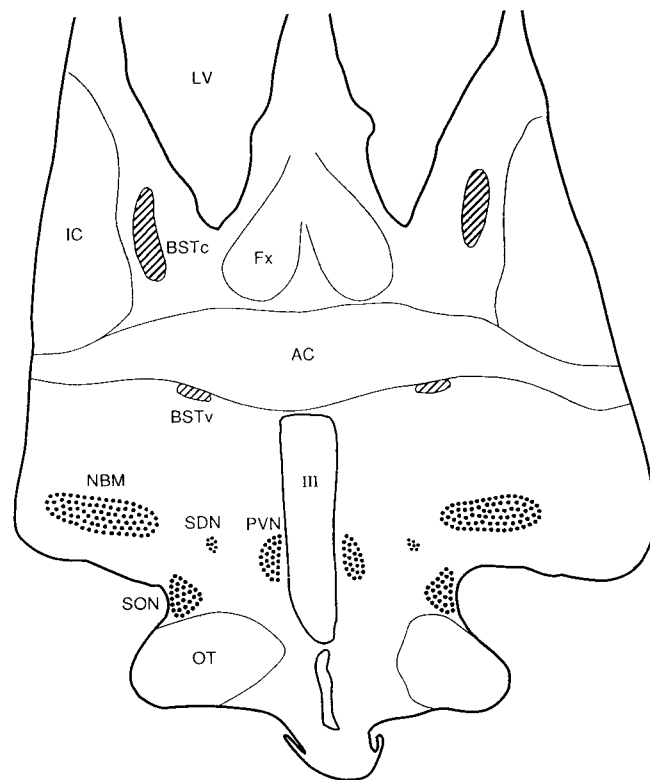


FIG. 1 Schematic frontal section through two subdivisions of the bed nucleus of the stria terminalis (BST). III, third ventricle; AC, anterior commissure; BSTc and BSTv, central and ventral subdivisions of the BST; Fx, fornix; IC, internal capsule; LV, lateral ventricle; NBM, nucleus basalis of Meynert; OT, optic tract; PVN, paraventricular nucleus; SDN, sexually dimorphic nucleus; SON, supraoptic nucleus.

We have studied the hypothalamus of six male-to-female transsexuals (T1–T6); this material was collected over the past eleven years. We searched for a brain structure that was sexually dimorphic but that was not influenced by sexual orientation, as male-to-female transsexuals may be 'oriented' to either sex with respect to sexual behaviour. Our earlier observations showed that the paraventricular nucleus (PVN), sexually dimorphic nucleus (SDN) and supraoptic nucleus (SON) did not meet these criteria (ref. 6 and unpublished data). Although there is no accepted animal model for gender-identity alteration, the bed nucleus of the stria terminalis (BST) turned out to be an appropriate candidate to study for the following reasons. First, it is known that the BST plays an essential part in rodent sexual behaviour^{3,4}. Not only have oestrogen and androgen receptors been found in the BST^{8,9}, it is also a major aromatization centre in the developing rat brain¹⁰. The BST in the rat receives projections mainly from the amygdala and provides a strong input in the preoptic-hypothalamic region^{11,12}. Reciprocal connections between hypothalamus, BST and amygdala are also well documented in experimental animals^{13–15}. In addition, sex differences in the size and cell number of the BST have been described in rodents which are influenced by gonadal steroids in development^{16–18}. Also, in humans a particular caudal part of the BST (BNST-dspm) has been reported to be 2.5 times larger in men than in women¹⁹.

The localization of the BST is shown in Fig. 1. The central part of the BST (BSTc) is characterized by its somatostatin cells and vasoactive intestinal polypeptide (VIP) innervation²⁰. We measured the volume of the BSTc on the basis of its VIP innervation (Fig. 2). The BSTc volume in heterosexual men ($2.49 \pm 0.16 \text{ mm}^3$) was 44% larger than in heterosexual women ($1.73 \pm 0.13 \text{ mm}^3$) ($P < 0.005$) (Fig. 3). The volume of the BSTc of heterosexual and homosexual men did not differ in any statist-

‡ To whom correspondence should be addressed.

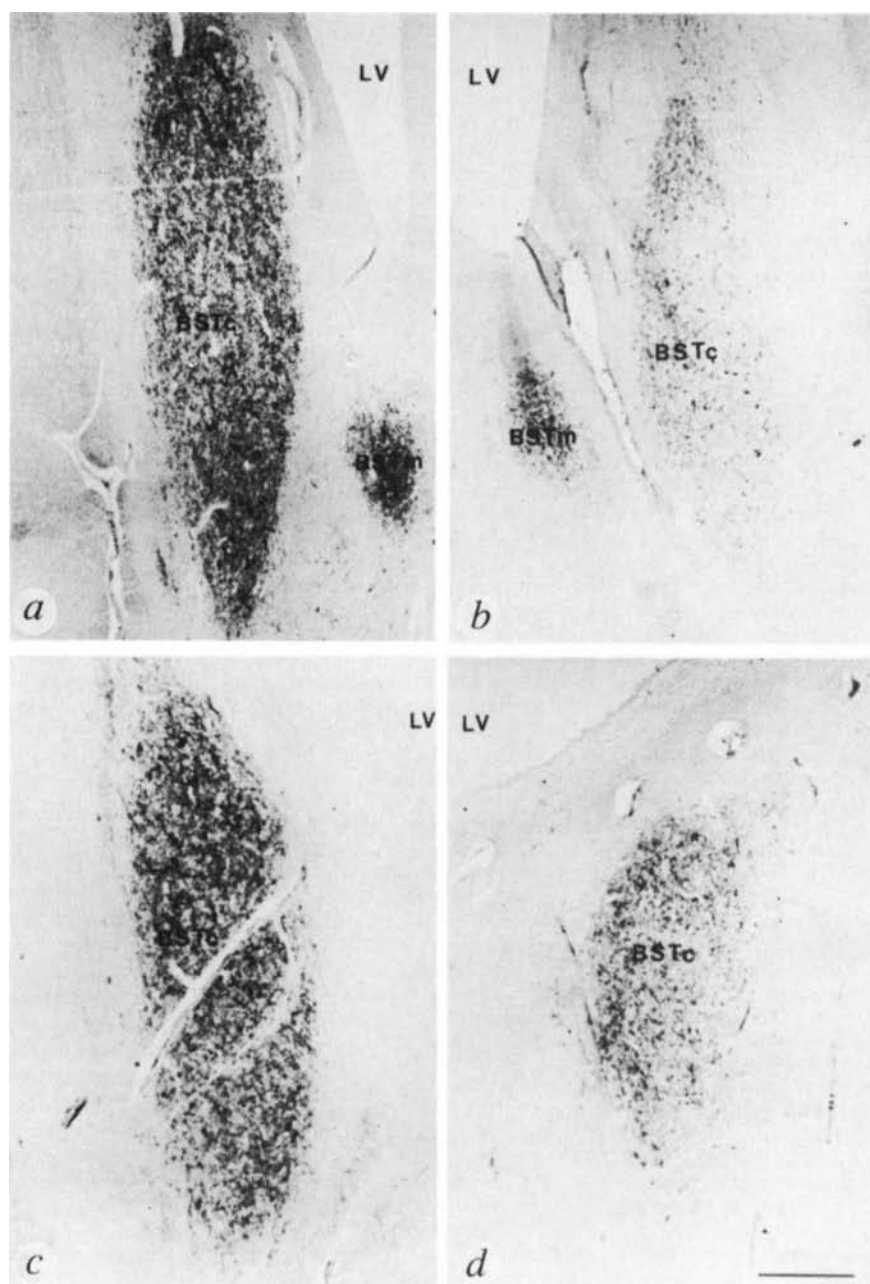


FIG. 2 Representative sections of the BSTc innervated by vasoactive intestinal polypeptide (VIP). a, Heterosexual man; b, heterosexual woman; c, homosexual man; d, male-to-female transsexual. Scale bar, 0.5 mm. LV, lateral ventricle. Note there are two parts of the BST in a and b: small medial subdivision (BSTm) and large oval-sized central subdivision (BSTc).

ically significant way ($2.81 \pm 0.20 \text{ mm}^3$) ($P=0.26$). The BSTc was 62% larger in homosexual men than in heterosexual women ($P<0.005$). AIDS did not seem to influence the size of the BSTc: the BSTc size of two heterosexual AIDS-infected women and three heterosexual AIDS-infected men remained well within the range of the corresponding reference group (Fig. 3). The AIDS-infected heterosexuals were therefore included in the corresponding reference group for statistical purposes. A small volume of the BSTc ($1.30 \pm 0.23 \text{ mm}^3$) was found in the male-to-female transsexuals (Fig. 3). Its size was only 52% of that found in the reference males ($P<0.005$) and 46% of the BSTc of homosexual males ($P<0.005$). Although the mean BSTc volume in the transsexuals was even smaller than that in the female group, the difference did not reach statistical significance ($P=0.13$). The volume of the BSTc was not related to age in any of the reference groups studied ($P>0.15$), indicating that the observed small size of the BSTc in transsexuals was not due to the fact that they were, on average, 10 to 13 years older than the hetero- and homosexual men.

The BST plays an essential role in masculine sexual behaviour and in the regulation of gonadotrophin release, as shown by studies in the rat^{3,4,21}. There has been no direct evidence that the BST has such a role in human sexual behaviour, but our demonstration of a sexually dimorphic pattern in the size of the human BSTc, which is in agreement with the previously described sex difference in a more caudal part of the human BST (BNST-dspm)¹⁹, indicates that this nucleus may also be involved in human sexual or reproductive functions. It has been proposed that neurochemical sex differences in the rat BST may be due to effects of sex hormones on the brain during development and in adulthood^{22,23}. Our data from humans however, indicate that BSTc volume is not affected by varying sex hormone levels in adulthood. The BSTc volume of a 46-year-old woman, who had suffered for at least one year from a tumour of the adrenal cortex that produced very high blood levels of androstenedione and testosterone, was within the range of that of other women (Fig. 3; S1). Furthermore, two postmenopausal women (aged over 70 years) showed a completely normal female-sized BSTc (Fig. 3;

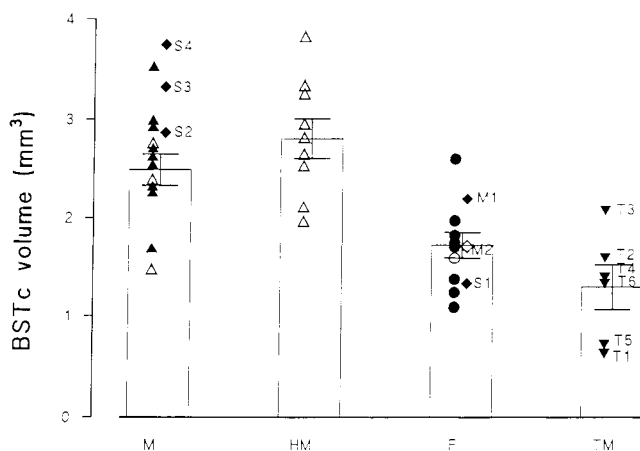


FIG. 3 Volume of the BSTc innervated by VIP fibres in presumed heterosexual males (M), homosexual males (HM), presumed heterosexual females (F) and male-to-female transsexuals (TM). The six transsexuals are numbered T1–T6. The patients with abnormal sex hormone levels are numbered S1–S4. M1 and M2, postmenopausal women. Bars indicate mean $\pm$ s.e.m. Open symbols: individuals who died of AIDS. METHODS. Brains of 42 subjects matched for age, postmortem time and duration of formalin fixation were investigated. The autopsy was performed after obtaining permission. For immunocytochemical staining of VIP, paraffin sections were hydrated and rinsed in Tris-buffered-saline TBS: 0.05 M Tris, 0.9% NaCl, pH 7.6). Sections were incubated with 200 μ l anti-VIP 1:1,000 in 0.5% Triton in TBS overnight at 4 °C. Immunocytochemical and morphometric procedures have been described^{25–27}. In brief, serial 6 μ m sections of the BSTc were studied by means of a digitizer (Calcomp 2000) connected to a HP-UX 9.0, using a Zeiss microscope equipped with a 2.5 $\times$ objective and with 10 $\times$ (PLAN) oculars. Staining was performed on every fiftieth section with anti-VIP. The rostral and caudal borders of the BSTc were assessed by staining every tenth section in the area. The volume of the BSTc was determined by integrating all the area measurements of the BSTc sections that were innervated by VIP fibres. In a pilot study, the size of the BSTc was measured on both sides in 8 subjects (5 females and 3 males) and no left–right asymmetries were observed: the left BSTc (1.71 ± 0.16 mm³) was comparable in size to that of the right BSTc (1.83 ± 0.30 mm³) ($P=0.79$). No asymmetry was observed in the BNST-dspm either¹⁹. The rest of our study was therefore performed on one side of the brain only. Brain weight of the male transsexuals ($1,385 \pm 78$ g) was not different from that of the reference males ($1,453 \pm 25$ g) ($P=0.61$) or that of the females ($1,256 \pm 35$ g) ($P=0.23$). The causes of death of the transsexuals were suicide (T1), cardiovascular disease (T2, T6), sarcoma (T3), AIDS, pneumonia, pericarditis (T4) and hepatic failure (T5). Sexual orientation of the subjects of the reference group (12 men and 11 women) was generally not known, but most of them were presumed heterosexual. Sexual orientation of 9 homosexuals was registered in the clinical records²⁸. Differences among the groups were tested two-tailed using the Mann–Whitney U -test. A 5% level of significance was used in all statistical tests.

M1, M2). As all the transsexuals had been treated with oestrogens, the reduced size of the BSTc could possibly have been due to the presence of high levels of oestrogen in the blood. Evidence against this comes from the fact that transsexuals T2 and T3 both showed a small, female-like BSTc (Fig. 3), although T2 stopped taking oestrogen about 15 months before death because her prolactin levels were too high and T3 stopped hormone treatment as a sarcoma was found about 3 months before death; also a 31-year-old man who suffered from a feminizing adrenal tumour which induced high blood levels of oestrogen, nevertheless had a very large BSTc (Fig. 3; S2).

Our results might also be explained if the female-sized BSTc in the transsexual group was due to a lack of androgens, because they had all been orchidectomized except for T4. We therefore studied two other men who had been orchidectomized because of cancer of the prostate (one and three months before death: S4 and S3, respectively), and found that their BSTc sizes were

at the high end of the normal male range. The BSTc size of the single transsexual who had not been orchidectomized (T4) ranged in the middle of the transsexual scores (Fig. 3). Not only were five of the transsexuals orchidectomized, they all used the antiandrogen compound cyproterone acetate (CPA). A CPA effect on the BSTc does not seem likely, because T6 had not taken CPA for the past 10 years, and T3 took no CPA during the 2 years before death and still had a female-sized BSTc.

In summary, our observations suggest that the small size of the BSTc in male-to-female transsexuals cannot be explained by differences in adult sex hormone levels, but is established during development by an organizing action of sex hormones, an idea supported by the fact that neonatal gonadectomy of male rats and androgenization of female rats induces significant changes in the number of neurons of the BST and suppresses its sexual dimorphism^{17,18}.

Considered together with information from animals, then our study supports the hypothesis that gender identity alterations may develop as a result of an altered interaction between the development of the brain and sex hormones^{3,6}. The direct action of genetic factors should also be considered on the basis of animal experiments²⁴.

We found no relationship between BSTc size and the sexual orientation of transsexuals, that is, whether they were male-oriented (T1, T6), female-oriented (T3, T2, T5) or both (T4). Furthermore, the size of the BSTc of heterosexual men and homosexual men did not differ, which reinforced the idea that the reduced BSTc size is independent of sexual orientation. In addition, there was no difference in BSTc size between early-onset (T2, T5, T6) and late-onset transsexuals (T1, T3), indicating that the decreased size is related to the gender identity alteration *per se* rather than to the age at which it becomes apparent. Interestingly, the very small BSTc in transsexuals appears to be a very local brain difference. We failed to observe similar changes in three other hypothalamic nuclei, namely PVN, SDN or SCN in the same individuals (unpublished data). This might be due to the fact that these nuclei do not all develop at the same time, or to a difference between these nuclei and the BST with respect to the presence of sex hormone receptors or aromatase. We are now studying the distribution of sex hormone receptors and the aromatase activity in various hypothalamic nuclei in relation to sexual orientation and gender. □

Received 23 May; accepted 28 August 1995.

1. Money, J. & Gaskin, R. J. *Int. J. Psychiatry* **9**, 249 (1970/71).
2. Gooren, L. J. G. *Psychoneuroendocrinology* **15**, 3–14 (1990).
3. Kawakami, M. & Kimura, F. *Endocrinol. Jap.* **21**, 125–130 (1974).
4. Emery, D. E. & Sachs, B. D. *Physiol. Behav.* **17**, 803–806 (1976).
5. Editorials *Lancet* **338**, 603–604 (1991).
6. Swaab, D. F. & Hofman, M. A. *Trends Neurosci.* **18**, 264–270 (1995).
7. Money, J., Schwartz, M. & Lewis, V. G. *Psychoneuroendocrinology* **9**, 405–414 (1984).
8. Sheridan, P. J. *Endocrinology* **104**, 130–136 (1979).
9. Commins, D. & Yahr, D. J. *comp. Neurol.* **231**, 473–489 (1985).
10. Jakab, R. L., Horvath, T. L., Leranth, C., Harada, N. & Naftolin, F. J. *Steroid Biochem. molec. Biol.* **44**, 481–498 (1993).
11. Eiden, E. L., Hökfelt, T., Brownstein, M. J. & Palkovits, M. *Neuroscience* **15**, 999–1013 (1985).
12. De Olmos, J. S. in *The Human Nervous System* (ed. Paxinos, G.) 597–710 (Academic, San Diego, 1990).
13. Woodhams, P. L., Roberts, G. W., Polak, J. M. & Crow, T. J. *Neuroscience* **8**, 677–703 (1983).
14. Simerly, R. B. *Trends Neurosci.* **13**, 104–110 (1990).
15. Aulison, M. et al. *Brain Res. Bull.* **34**, 319–337 (1994).
16. Bleier, R., Byne, W. & Siggekow, I. J. *comp. Neurol.* **212**, 118–130 (1982).
17. Del Abril, A., Segovia, S. & Guillamón, A. *Dev. Brain Res.* **32**, 295–300 (1987).
18. Guillamón, A., Segovia, S. & Del Abril, A. *Dev. Brain Res.* **44**, 281–290 (1988).
19. Allen, L. A. & Gorski, R. A. J. *comp. Neurol.* **302**, 697–706 (1990).
20. Walter, A., Mai, J. K., Lanta, L. & Görcs, T. J. *Chem. Neuroanat.* **4**, 281–298 (1991).
21. Claro, F., Segovia, S., Guillamón, A. & Del Abril, A. *Brain Res. Bull.* **36**, 1–10 (1995).
22. Simerly, R. B. & Swanson, L. W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2087–2091 (1987).
23. De Vries, G. J. N. *Neuroendocr.* **20**, 1–13 (1990).
24. Pilgrim, C. & Reiser, I. *Horm. metab. Res.* **24**, 353–359 (1992).
25. Swaab, D. F., Zhou, J. N., Ehlhart, T. & Hofman, M. A. *Brain Res.* **79**, 249–259 (1994).
26. Zhou, J. N., Hofman, M. A. & Swaab, D. F. *Neurobiol. Aging* **16**, 571–576 (1995).
27. Zhou, J. N., Hofman, M. A. & Swaab, D. F. *Brain Res.* **672**, 285–288 (1995).
28. Swaab, D. F. & Hofman, M. A. *Brain Res.* **537**, 141–148 (1990).

ACKNOWLEDGEMENTS. We thank B. Fisser, H. Stoffels, G.v.d. Meulen, T. Eikelboom and W. T. P. Verweij for their help, and R. M. Buijs, M. A. Corner, E. Fliers, A. Walter and F. W. van Leeuwen for their comments. Brain material was provided by the Netherlands Brain Bank (coordinator R. Ravid). This work was supported by NWO.

Remodelling of hand representation in adult cortex determined by timing of tactile stimulation

Xiaoqin Wang*, Michael M. Merzenich, Koichi Sameshima* & William M. Jenkins

Coleman Laboratory and Keck Center for Integrative Neuroscience, University of California at San Francisco, Box 0732, San Francisco, California 94143, USA

THE primate somatosensory cortex, which processes tactile stimuli, contains a topographic representation of the signals it receives, but the way in which such maps are maintained is poorly understood. Previous studies of cortical plasticity^{1–20} indicated that changes in cortical representation during learning arise largely as a result of hebbian synaptic change mechanisms. Here we show, using adult owl monkeys trained to respond to specific stimulus sequence events, that serial application of stimuli to the fingers results in changes to the neuronal response specificity and maps of the hand surfaces in the true primary somatosensory cortical field (S1 area 3b). In this representational remodelling stimuli applied synchronously to the fingers resulted in these fingers being integrated in their representation, whereas fingers to which stimuli were applied asynchronously were segregated in their representation. Ventro-posterior thalamus response maps derived in these monkeys were not equivalently reorganized. This representational plasticity appears to be cortical in origin.

The phenomenology of cortical representational change in learning arises, to a large extent, by the operation of hebbian, that is, coincident input-dependent, synaptic change mechanisms^{7,21}. Inputs that nearly simultaneously engage cortical neurons are integrated in cortical neuronal responses. In one class of studies that argued for a role of input timing in establishing cortical 'map' topologies, digits were surgically fused⁷, or long-fused digits were separated¹⁵, thereby altering the detailed structures of temporally coincident and non-coincident inputs delivered from the hand into the somatosensory nervous system. The former manipulation resulted in a breakdown of the normally sharply segregated representations of adjacent digits in the area 3b hand zone⁷; the latter manipulation resulted in an apparent reseparation of fused digit representations in the S1 cortex¹⁵. Taken together, these studies indicated that the normally continuous representations of individual digits and the normally discontinuous representations of different, adjacent digital surfaces both arose dynamically from the operation of cortical plasticity mechanisms. They inferred that, in addition to hebbian coincident input-based response remodelling, there may be a second important principle for cortical representational remodelling. By the operation of cortical plasticity mechanisms in learning, consistently non-coincident inputs may be actively segregated from one other in their distributed cortical representations.

To test these hypotheses further and to define the specific temporal input conditions under which inputs are integrated (or, hypothetically, actively segregated) by cortical plasticity mechanisms, we trained unrestrained adult owl monkeys to place one hand on a form-fitting handgrip to discriminate tactile stimulus sequences delivered to their hands by two narrow bars. One bar simultaneously stimulated a narrow line of skin crossing the distal segments (phalanges) of digits 2, 3 and 4; the second bar excited a narrow line of skin crossing the middle or (in other

monkeys) the proximal segments of the same three digits (Fig. 1). A series of ramped step indentations (50 ms duration) with interstimulus intervals of 200–300 ms were alternately applied to the skin by the two bars as background tactile stimulation (Fig. 1b). The target signal for a particular monkey was either two consecutive step indentations applied to one or the other bar (Fig. 1b), or a 50–60 Hz sinusoidal vibration substituted for a step indentation on one or the other bar. Both background and target stimuli had the same fixed amplitude (50–100 μ m). In all cases, the target signal appeared randomly and with equal probability on either of the two bars. All monkeys attained a highly stereotyped hand position on the handgrip, insuring that precisely the same skin surfaces were stimulated on each behavioural trial. The training created a heavy schedule of stereotyped, temporally coincident inputs across fingertips and fingerbases. However, distal versus proximal digit segment surfaces were consistently temporally non-coincidentally stimulated. For primates and humans, the usual stimulation on several fingers generated by normal use of hands is in fact substantially and consistently non-simultaneous, regarding the 5–20 ms integrative time constant that appears to govern cortical coincident input-based plasticity. By contrast, here we consistently delivered cross-finger stimuli that were absolutely coincident. Along the length of a finger in the normal case, immediately adjacent skin loci are excited nearly coincidentally in time on a heavy, natural schedule, a situation that does not apply to the stimulation of adjacent fingers. In our experimental cases, adjacent skin along the length of a finger, in sharp contra-distinction to the natural case, is consistently stimulated non-coincidentally.

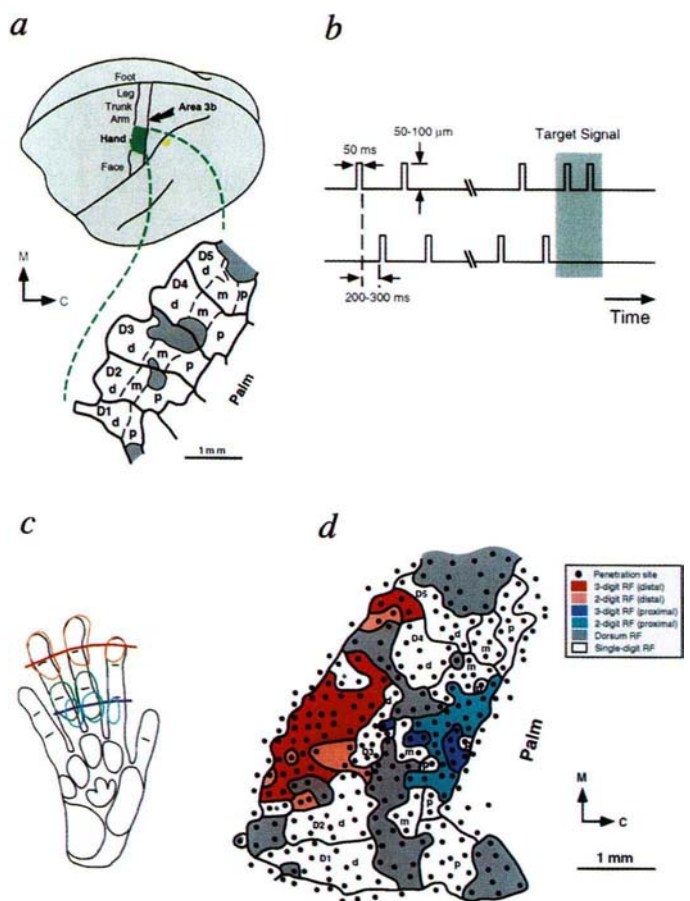
After a monkey learned to position its hand with high reliability, it underwent 4–6 weeks of intensive training at this task, performing several hundred behavioural trials each day. Immediately after this training period, we conducted a terminal electrophysiological neuronal response mapping study in which we defined in detail how the surfaces of the hand were represented in cortical layer 3 in area 3b in the contralateral cerebral hemisphere (Fig. 1d). In striking contra-distinction to normal monkeys, many sampled cortical units in the hand sector of cortical area 3b in these monkeys had multiple-digit receptive fields (mRFs). Such mRF neurons had individually definable receptive fields on two or on all three of the adjacent, trained digit tips, or digit bases (Figs 1c and 2A). In equivalent maps derived in area 3b in normal, untrained owl monkeys (Fig. 1a), two-digit RFs were rarely recorded; three-digit RFs in the centre of the hand zone of area 3b were never seen in controls.

Reorganized cortical maps in these monkeys were characterized by two large, continuous zones in which all sampled units had mRFs. In the representative case illustrated in Fig. 1d, all neurons in the red and blue zones had mRFs on the distal and proximal-middle finger segments, respectively. Within these large cortical regions, mRFs recorded at separated cortical locations commonly extensively overlapped with one another (Fig. 3b, d, top left panels), and the normally segregated representation of adjacent digits was completely broken down. Of 102, 100 and 127 neuronal unit samples recorded in the area 3b zone of 'trained' digits 2, 3 and 4 in these three experimental monkeys, 44%, 35% and 56% of studied units had mRFs. Adjacent to each zone in which all neurons had mRFs were area 3b sectors in which units had single-digit RFs, albeit with still-abnormal representational topographies.

For mRF units, evoked neuronal spike discharges were typically almost as strong and occurred with equivalently short latencies when any one of the individual RF components of mRF fields located on two or three trained fingers was stimulated separately (Fig. 2A, B). The mRF neurons were never excited significantly by stimulation of adjacent, untrained digits (D1 and 5; Fig. 2A). The component receptive fields comprising mRFs recorded in trained monkeys exhibited consistently low thresholds to cutaneous skin stimulation and, in contra-distinction to controls (Fig. 2C, bottom left), had reliably

* Present addresses: Department of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Avenue, Traylor 710, Baltimore, Maryland 21205, USA (X.W.); University of Sao Paulo, School of Medicine, Discipline of Medical Informatics, Avenida Dr Arnaldo 455, 01246-903, Sao Paulo, Brazil (K.S.).

FIG. 1 *a*, Area 3b in the primary somatosensory cortex (S1) of adult owl monkeys (*Aotus nancymai*) contains a single, orderly representation of the hand surfaces, with sharp discontinuities separating the representations of individual fingers^{22–25}. A representative cortical map of a hand in area 3b (bottom) of the contralateral hemisphere (top) obtained from a normal, untrained adult owl monkey is shown. Maps defined in individual monkeys share common features and vary in representational details²². Cortical zones with cutaneous responses evoked from different ventral glabrous hand surfaces are outlined and marked by letters. Cortical zones representing the dorsal surfaces of the hand are indicated by grey shading. In normal monkeys, area 3b glabrous RFs recorded at given cortical sites were almost always located on single digits. D1–D5 refer to cortical zones in which all neurons responded only to stimulation of the glabrous skin surfaces of digits 1–5; d, m and p refer to distal, middle and proximal finger segments, respectively. *b*, An example of a tactile stimulus sequence used in behavioural training is shown, with fingertip and fingerbase stimulation represented by upper and lower event tracks. The target stimulus is marked by the grey shading. The stimulus amplitude and interstimulus intervals (ISIs) had a fixed value for each monkey (see text). For the monkey whose area 3b map is shown in *d*, each individual step indentation was 75 μm in amplitude, with an ISI of 300 ms. *c*, Three examples of representative mRFS recorded in area 3b in a trained monkey. As is typical for mRFS, all RF components included skin surfaces that were stimulated in training by one (not both) of the two stimulators. Locations of the two narrow tactile stimulation bars striking the hand are indicated by dark red (distal) and dark blue (proximal) lines. *d*, A typical reorganized map of the hand representation recorded in cortical area 3b of a trained adult owl monkey. Digits 2, 3 and 4 of the right hand were engaged in behavioural training. This map was constructed on the basis of cortical units recorded at 300 locations, marked by dots in the figure, sampled with a density of 30–50 samples per mm^2 in cortical layer 3. Cortical zones representing individual digits are outlined. All sampled cortical units in the red zones had 2 (light red) or 3 (dark red) RF components on distal digit segments; those in blue zones had 2 (light blue) or 3 (dark blue) RF components on proximal/middle digit phalanges. Zones in which sampled units had RFs on the dorsal hand surfaces are shaded in grey. Cortical units in white zones had single-digit ventral glabrous RFs. The asterisks indicate the zones with Pacinian-like responses. In this animal, 127 sampled cortical units had their RFs on the glabrous skin surfaces of the trained digits (D2, 3 and 4); 71 of them (56%) had mRFS. **METHODS.** Three adult owl monkeys were trained in the two-bar discrimination task. Tactile stimuli were delivered by two calibrated, displacement feedback-controlled mechanical stimulators. The behavioural training setup and paradigm were similar to those reported previously²⁶. Each monkey performed 300–500 trials per day, 6 or 7 days per week. Progress in training was marked by increasing success rates and decreasing behavioural response latencies and variances. The task was difficult for these monkeys, but high performance levels were ultimately achieved for each animal (>80–90% correct). In physiological mapping experiments, single or multi-unit responses were recorded from layer 3 of cortical field 3b in the hemisphere contralateral to the trained hand under barbiturate anaesthesia which was maintained through intravenous injections of sodium pentobarbital (1–5 mg h^{-1}) throughout these 3–5-day-long experiments. Recordings were made from parallel penetrations introduced perpendicular to the cortical surfaces, using parylene-coated tungsten microelectrodes with impedance of 1–2 $\text{M}\Omega$ at 1 kHz. Each penetration was marked on a computer-captured magnified ($\times 40$) video image of the cortical surface, with reference to vascular details. Cutaneous RFs were defined as those skin surfaces evoking a reliable suprathreshold neural response with



application of just-visible tactile stimulation with fine opaque glass probes (tip diameter $< 1 \text{ mm}$). They were drawn on magnified video hand images captured and stored by computer. Many RFs were also defined with the use of von Frey hairs, which delivered indenting stimuli of nearly constant force in each application. Under these anaesthetic conditions, the location and size of RFs of area 3b neurons have been found to be very similar to those recorded in unanaesthetized owl monkeys²⁷. After cortical RF mapping, neuronal discharges were recorded in a subset of studied cortical sites, with controlled tactile stimuli applied to relevant skin surfaces. In these cases, the hand of the monkey was stabilized for the delivery of calibrated indenting stimuli using displacement feedback-controlled mechanical stimulators. After derivation of these quantitative cortical data, electrode tracks were advanced into the ventroposterior thalamus contralateral to the trained hand, and samples of unit responses were derived from control and experimental fingers in this mainline somatosensory nucleus. At the completion of physiological recordings, electrolytic lesions were made and histology subsequently performed to objectively define the borders of cortical area 3b, and to reconstruct ventroposterior thalamus mapping microelectrode penetration tracks.

driven responses evoked from multiple digits across a wide range of stimulus intensities (Fig. 2C, top left). Thus mRFS appeared to be truly integrating responses from consistently simultaneously struck multiple digital surfaces in area 3b of the cortex.

RFs recorded in topographically reorganized mRF and single-digit RF zones typically included either multiple- (or single-) digit distal or proximal trained skin sites, but not both (Fig. 3). In contradiction to the normal shifted-overlap receptive field progressions recorded in area 3b in normal monkeys, receptive fields recorded at different sites in mRF zones were often virtually identical for neuronal samples separated by long distances across area 3b.

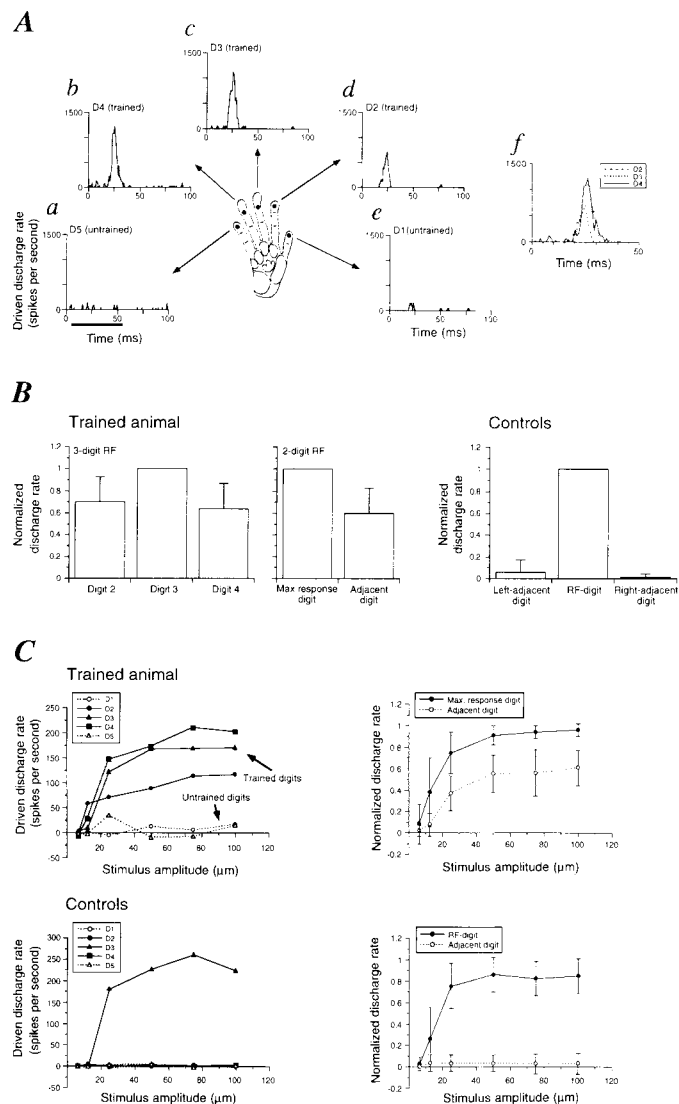
The two sets of mRFS contained distal or proximal trained skin sites actually abutted each other on the hand surface (Fig. 3b, d, top panels). However, the cortical zones representing these two sets of mRFS were sharply separated topographically in area 3b, in general by more than $\sim 1 \text{ mm}$ (Fig. 3a, c). In addition to the breakdown of the representational discontinuities that normally separate digit representations in area 3b in monkeys, a prominent area 3b band emerged in trained monkeys, in which all units responded exclusively to dorsal skin inputs. This dorsal skin representational band was located between the zones of representation of consistently non-simultaneously excited distal and proximal skin surfaces on the trained fingers. The emergence of this hand dorsum zone (Fig. 3a, c, black areas) further exag-

FIG. 2 Quantitative evaluations of training induced cortical response changes in area 3b. In A–C, tactile stimuli were ramped step indentations (amplitude 50 μm , duration 50 ms, onset at 5 ms) similar to those used in training, but here applied separately to each finger by a mechanical stimulator. Neuronal responses were recorded for 500–1,000 ms for 20 stimulus repetitions. Driven discharge rates (stimulus-evoked discharge rate minus spontaneous discharge rate) were computed from a window 5–45 ms after stimulus onset. **A**, **a–e**, Post stimulus histograms (PSTHs) of stimulus-evoked responses recorded from a typical mRF unit in a trained monkey (binwidth, 1 ms). This typical mRF cortical unit had RF components on the distal segment of digits 2, 3 and 4. In **f**, PSTHs of the three trained digits (**b–d**) are shown together on an expanded time scale to illustrate their nearly identical response latencies and similar profiles. **B**, for trained animals, averaged cortical responses evoked by tactile stimulation on adjacent digits are shown for all distal mRF units studied quantitatively in two trained monkeys. For 3-digit mRF units (left), driven discharge rates of digits 2, 3 and 4 were normalized by that of digit 3 for each unit. For 2-digit mRF units (right), the digit that gave the strongest response was designated as the 'maximum-response digit'. Driven discharge rates of adjacent digits were normalized by that of the maximum response digit for each unit. For 3-digit mRF units ($N=9$), 0.700 ± 0.228 (s.d.) (digit 2), 1.0 ± 0 (digit 3) and 0.634 ± 0.230 (digit 4); 2-digit mRF units ($N=7$), 1.0 ± 0 (maximum-response digit) and 0.602 ± 0.227 (adjacent digit). For controls, averaged cortical responses evoked by tactile stimulation on adjacent digits recorded in a control hemisphere representing an untrained hand are shown for all quantitatively studied cortical units in this animal whose RFs were on the distal phalanges of digits 2, 3 and 4. For each unit, driven discharge rates of adjacent digits were normalized by that of the digit on which a RF was defined (the RF-digit) ($N=17$): 0.057 ± 0.112 (left-adjacent digit), 1.0 ± 0 (RF-digit) and 0.013 ± 0.030 (right-adjacent digit). **C**, For trained animals: left, driven discharge rates versus stimulus amplitude for the typical mRF unit shown in **A**. The behavioural training stimuli in this monkey had a fixed amplitude of 75 μm . Filled symbols, trained digits (D2, 3 and 4); Open symbols, untrained digits (D1 and 5). Right, averaged normalized discharge rates versus stimulus amplitude for all quantitatively studied mRF units ($N=9$, seven 3-digit mRF units, two two-digit mRF units) with RFs on the distal segments in one trained monkey (vertical bars, s.d.). For each unit, responses at different stimulus amplitudes and skin sites were normalized by the maximum driven discharge rate obtained for that unit. The responses from the digits giving the maximum driven discharge rate are shown with filled symbols, those from adjacent digit(s) with open symbols ($N=16$, one adjacent digit for each 2-digit mRF unit, two adjacent digits for each 3-digit mRF unit). For controls: left, driven discharge rates versus stimulus amplitude for a typical cortical unit recorded in an untrained control hemisphere. This unit had a RF on the distal segment of digit 3. Right, averaged normalized discharge rates versus stimulus amplitude for all recorded units ($N=17$) with RF on distal phalanges of digits 2, 3 or 4 in one untrained animal (vertical bars, s.d.). For each unit, responses at different stimulus amplitudes and skin sites were normalized by the maximum driven discharge rate

generated the representational segregation of skin surfaces consistently struck non-simultaneously. No equivalent mid-finger dorsal RF zone dividing the representations of these three fingers was recorded in cortical area 3b in normal untrained owl monkeys, in which dorsal hand surfaces are usually represented in patches located between finger representations, and/or along the lateral or medial margins of the area 3b hand representation²² (Fig. 1a).

Although this particular behavioural training experience altered the response properties of most sampled units in the digital representational sector of the hand zone of area 3b, the skin surfaces of the trained fingers were still represented (albeit in a clearly distorted and incomplete way in comparison with normal controls) by cortical units that had single-digit RFs (Fig. 3b, d, bottom right).

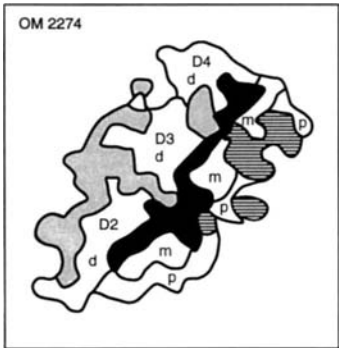
An important question concerning this learning-induced cortical representational remodelling is whether these plastic changes were specifically due to a cortical reorganization or merely indicated subcortical plastic changes. We examined this directly by recording responses from neurons in the hand representational region in the lateral division of the ventroposterior



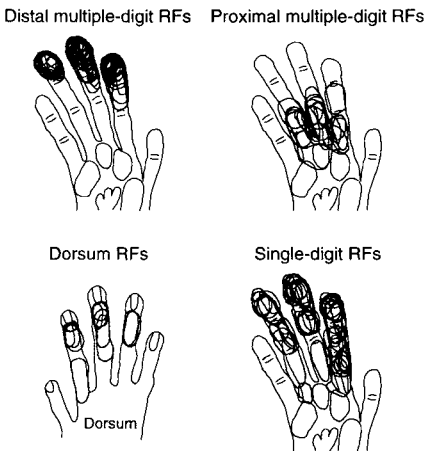
obtained for that unit. The responses from RF-digits are shown with filled symbols ($N=17$), those from adjacent digits with open symbols ($N=34$, two adjacent digits for each unit).

nucleus of the thalamus (VPL). In striking contrast to area 3b, RFs of all sampled VPL neurons representing digits 2, 3 and 4 were restricted to the surfaces of a single digit, as is generally the case in untrained owl monkeys, and there was no evidence of the representational segregation or a dorsal hand representational band separating consistently non-simultaneously excited ventral glabrous hand surfaces. Representative thalamic RF sequences from 'trained' and control digit representational zones are shown for a typical monkey in Fig. 4. Of area 3b units representing the trained digits, 56% had mRFs. Thus the emergence of integrated mRFs and of training-induced representational segregation recorded in cortical area 3b had no parallels in its principal thalamic source nucleus, the VPL; these main features of cortical representational remodelling arising from this behaviourally controlled hand use appear to be cortical in origin. Conclusions from this primate study are in accord with experience-induced remodelling in rats^{16, 18}, in which the plastic changes in the barrel cortex have been found to be induced by whisker pairing. In that model, as in these monkey studies, behaviourally induced changes were believed to be primarily cortical in origin¹⁷.

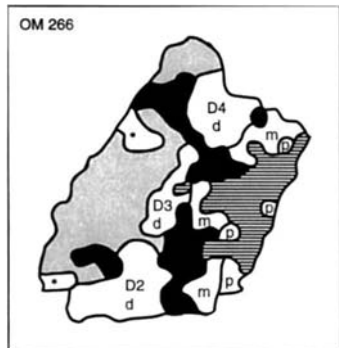
a Cortical maps of trained digit surfaces



b Receptive fields in different area 3b zones



c Cortical maps of trained digit surfaces



d Receptive fields in different area 3b zones

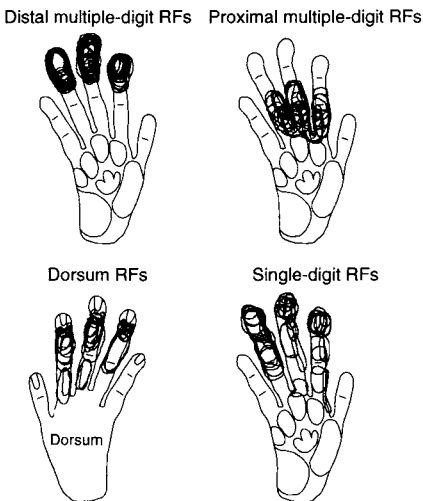


FIG. 3 Cortical zones representing different skin surfaces on the three trained fingers (digits 2, 3 and 4), partitioned using four RF categories, are shown for two monkeys (*a* and *c*). The corresponding RFs are shown in *b* and *d*, respectively. Distal multiple-digit RFs (*a* and *c*, grey shading; *b* and *d*, top left). Proximal multiple-digit RFs (*a* and *c*, horizontal striping; *b* and *d*, top right). Dorsal (hairy) RFs (*a* and *c*, black shading; *b* and *d*, bottom left). Single-digit RFs (*a* and *c*, white areas; *b* and *d*, bottom right). The asterisks indicate the zones with Pacinian-like responses.

Representative VPL (somatosensory thalamus) responses

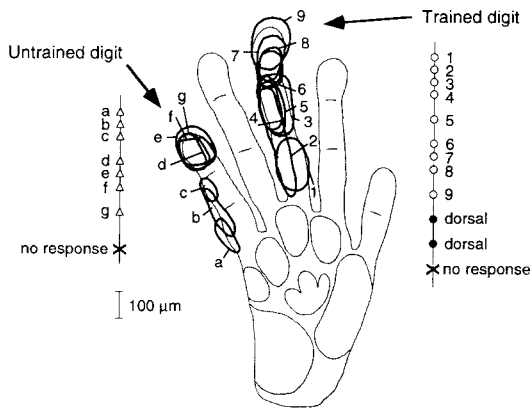


FIG. 4 Examples of RFs from two representative microelectrode recording tracks crossing trained and untrained digit representational regions of the lateral part of the ventroposterior nucleus of the thalamus (VPL). Thalamic RFs recorded from these two tracks in this monkey, whose reorganized area 3b map is shown in Fig. 1*d*, were on digit 5 (an untrained finger) and digit 3 (a trained finger), respectively. Only single-digit RFs, and no segregation of different digital surfaces by a dorsal representational band, were recorded in trained-digit representational zones, in striking contrast to cortical findings. These microelectrode tracks, representative of others recorded in the experiment, were reconstructed histologically; both spanned the dorsoventral extent of VPL.

VPL experiment summary:
99 units sampled
64 on trained digits (D2,3,4), 100% were single-digit RF
13 on untrained digits (D1,5), 100% were single-digit RF
22 on palm or dorsum

In the 'new' hand map generated in these monkeys by this new occupational hand use, hand surfaces that are normally sharply segregated in their representation in area 3b came to be integrated in representation, whereas digital surfaces that are normally represented continuously by shifted-overlap receptive-field progressions came to be sharply segregated from each other. These changes apparently resulted from the fact that the heavy schedules of input generated by this training were spatiotemporally orthogonal to those generated by normal hand use. These findings provide further direct evidence that cortical neuronal response specificity, and the cortical maps that derive from it, are established dynamically by the plastic operations of the cerebral cortex. They provide another clear demonstration that cortical response integration as defined by the cutaneous receptive field is coincident-input based. They also demonstrate the active network processes that are responsible for non-coincident input-based segregation, hypothetically underlying stimulus differentiation, recognition and categorization. □

Received 23 May; accepted 30 August 1995.

1. Kaijaska, J. & Pomeranz, B. *J. Neurophysiol.* **42**, 618–633 (1979).
2. Rasmussen, D. D. *J. comp. Neurol.* **205**, 313–326 (1982).
3. Merzenich, M. M. *et al. Neuroscience* **8**, 33–55 (1983).
4. Merzenich, M. M. *et al. Neuroscience* **10**, 639–665 (1983).
5. Wall, J. T., Felleman, D. J. & Kaas, J. H. *Science* **221**, 771–773 (1983).

6. Merzenich, M. M. *et al. J. comp. Neurol.* **224**, 591–605 (1984).
7. Clark, S. A., Allard, T., Jenkins, W. M. & Merzenich, M. M. *Nature* **332**, 444–445 (1988).
8. Calford, M. B. & Tweedale, R. *Nature* **332**, 446–448 (1988).
9. Jenkins, W. M., Merzenich, M. M., Ochs, M. T., Allard, T. & Guic-Robles, E. *J. Neurophysiol.* **63**, 82–104 (1990).
10. Pons, T. P. *et al. Science* **252**, 1857–1860 (1991).
11. Garraghty, P. E. & Kaas, J. H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6976–6980 (1991).
12. Recanzone, G. H., Merzenich, M. M., Jenkins, W. M., Grajski, K. A. & Dinse, H. R. *J. Neurophysiol.* **67**, 1031–1056 (1992).
13. Code, R. A., Eslin, D. E. & Juliano, S. L. *Expl Brain Res.* **88**, 341–344 (1992).
14. Pascual-Leone, A. & Torres, F. *Brain* **116**, 39–52 (1993).
15. Mogilner, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **90**, 3593–3597 (1993).
16. Diamond, M. E., Armstrong-James, M. & Ebner, F. F. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2082–2086 (1993).
17. Diamond, M. E., Huang, W. & Ebner, F. F. *Science* **265**, 1885–1888 (1994).
18. Armstrong-James, M., Diamond, M. E. & Ebner, F. F. *J. Neurosci.* **14**, 6978–6991 (1994).
19. Yang, T. T. *et al. Neuroreport* **5**, 701–704 (1994).
20. Xerri, C., Stern, J. M. & Merzenich, M. M. *J. Neurosci.* **14**, 1710–1721 (1994).
21. Merzenich, M. M., Grajski, K. A., Jenkins, W. M., Recanzone, G. H. & Peterson, B. *Cold Spring Harb. Symp. quant. Biol.* **55**, 873–887 (1991).
22. Merzenich, M. M. *et al. J. comp. Neurol.* **258**, 281–296 (1987).
23. Merzenich, M. M., Kaas, J. H., Sur, M. & Lin, C. S. *J. comp. Neurol.* **181**, 41–73 (1978).
24. Sur, M., Merzenich, M. M. & Kaas, J. H. *J. Neurophysiol.* **44**, 295–311 (1980).
25. Kaas, J. H., Merzenich, M. M. & Killackey, H. P. A. *Rev. Neurosci.* **6**, 325–356 (1983).
26. Recanzone, G. H., Jenkins, W. M., Hradek, G. T. & Merzenich, M. M. *Behav. Res. Meth. Instrum. Comput.* **23**, 357–369 (1991).
27. Stryker, M. P., Jenkins, W. M. & Merzenich, M. M. *J. comp. Neurol.* **258**, 297–303 (1987).

ACKNOWLEDGEMENTS. We thank W. Bruszewski, B. Peterson and M. Fong for technical support; P. Ralston, D. Ralston and E. S. Ruthazer for histological work; D. Buonomano and C. deCharms for comments on the manuscript; and P. H. Bedenbaugh, R. Beitel and B. Peterson for help with electrophysiological experiments. This work was supported by a Kleberg Foundation Postdoctoral Fellowship (to X.W.), a FAPESP grant (to K.S.), and an NIH grant (to M.M.M.), the Coleman Fund and HRI.

Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons

S. R. Cobb*, E. H. Buhl*, K. Halasy*†, O. Paulsen* & P. Somogyi*

* MRC Anatomical Neuropharmacology Unit, University Department of Pharmacology, Mansfield Road, Oxford OX1 3TH, UK

† Department of Zoology & Cell Biology, Jozsef Attila University, Szeged H-6722, Hungary

SYNCHRONIZATION of neuronal activity is fundamental in the operation of cortical networks¹. With respect to an ongoing synchronized oscillation, the precise timing of action potentials is an attractive candidate mechanism for information coding^{2–5}. Networks of inhibitory interneurons have been proposed to have a role in entraining cortical, synchronized 40-Hz activity^{6,7}. Here we demonstrate that individual GABAergic interneurons⁸ can effectively phase spontaneous firing and subthreshold oscillations in hippocampal pyramidal cells at θ frequencies (4–7 Hz). The efficiency of this entrainment is due to interaction of GABA_A-receptor-mediated hyperpolarizing synaptic events with intrinsic oscillatory mechanisms tuned to this frequency range in pyramidal cells. Moreover, this GABAergic mechanism is sufficient to synchronize the firing of pyramidal cells. Thus, owing to the divergence of each GABAergic interneuron^{9,10}, more than a thousand pyramidal cells may share a common temporal reference established by an individual interneuron.

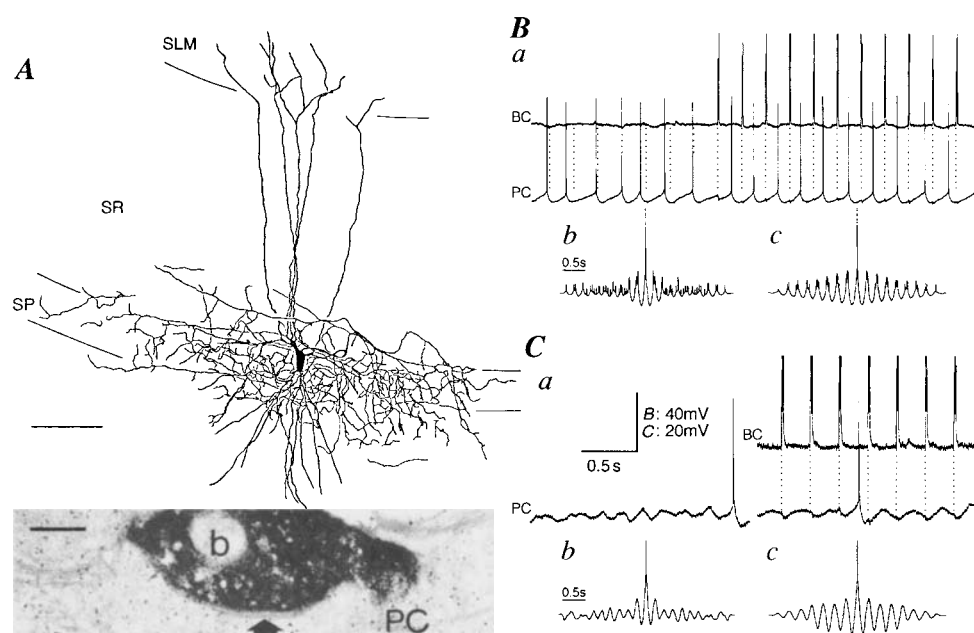
Cortical oscillatory activity at the electroencephalogram (EEG) level¹¹ has been correlated to distinct patterns of behaviour. One form of rhythmic activity, θ oscillations (4–7 Hz), are prominent in the hippocampal EEG during exploratory behaviour¹² and have been proposed to serve as a reference for coding by 'place cells'¹³. At the cellular level, intrinsic membrane conductances support θ -frequency membrane-potential oscillations in individual CA1 pyramidal cells¹³. Moreover, intracellu-

lar recordings from pyramidal cells during θ activity have revealed rhythmic chloride-mediated conductances that originate close to the cell body^{14,15}, suggesting that perisomatic GABA_A-receptor-mediated synaptic events contribute during intracellular θ oscillatory activity. We have tested, in the absence of patterned excitatory input, whether rhythmic activation of individual GABAergic interneurons, which selectively innervate the perisomatic region of pyramidal cells, can synchronize θ -frequency oscillations in CA1 pyramidal cells in hippocampal slices.

Anatomically identified GABAergic interneurons were either basket cells ($n = 8$; Fig. 1A) making synapses mainly on somata ($53 \pm 15\%$) and proximal dendrites ($44 \pm 15\%$), or axo-axonic cells ($n = 2$) terminating on axon initial segments of pyramidal cells (100%). Both basket and axo-axonic cells elicited GABA_A-receptor-mediated inhibitory postsynaptic potentials (i.p.s.ps)⁸ in simultaneously recorded pyramidal cells ($n = 17$). Rhythmic activation of presynaptic basket and axo-axonic interneurons at 1–8 Hz instantly phase-locked firing in simultaneously recorded postsynaptic pyramidal cells ($n = 7$ of 7 cells tested; Fig. 1B). By entraining the pyramidal cell, the interneuron could either decrease or increase the firing rate of the pyramidal cell. During entrainment, interneurons and pyramidal cells fired at alternate phases, as occurs during θ activity *in vivo*¹⁶. Subthreshold membrane-potential oscillations were also entrained by rhythmic activation of single basket and axo-axonic cells ($n = 5$ of 5; Fig. 1C), indicating that interneurons can entrain pyramidal cell activities both at sub- and suprathreshold levels.

To elucidate the mechanism underlying inhibitory phasing of pyramidal neurons, unitary i.p.s.ps were evoked at different membrane potentials. At depolarized membrane potentials at which firing occurred sporadically in the pyramidal cell, single spikes or short trains of action potentials in the inhibitory interneuron produced a hyperpolarizing i.p.s.p. in the pyramidal neuron followed by 'rebound' action potentials^{8,17} ($n = 10$ of 13; Fig. 2A, a). These action potentials preferentially occurred in a time window corresponding to the depolarizing overshoot following i.p.s.ps at subthreshold membrane potentials (Fig. 2A, b). Brief hyperpolarizing current pulses mimicking i.p.s.ps delivered by the somatic recording electrode produced a similar depolarizing overshoot ($n = 8$; Fig. 2B), indicating that the rebound was due to the i.p.s.p.-associated hyperpolarization interacting with

FIG. 1 Phase-locking of pyramidal cell discharge by a single presynaptic basket cell. **A**, Light-microscope-based reconstruction of a hippocampal basket cell (BC) with extensive axon mainly restricted to the cell body layer of area CA1 where it forms type II (symmetrical) synaptic contacts with somata (64%) and proximal dendrites (36%) of pyramidal cells. SP, stratum pyramidale; SR, stratum radiatum; SLM, stratum lacunosum-moleculare. Scale bar, 100 μ m. Bottom: electron micrograph of a symmetrical synapse (arrow) between a terminal (b) of the labelled basket cell and an unlabelled pyramidal cell (PC) soma. Scale bar, 0.2 μ m. Action potentials evoked by brief depolarizing current pulses to this basket cell produced fast i.p.s.ps in a simultaneously recorded pyramidal cell (not shown). The reversal potential was -69 mV and the average i.p.s.p. conductance was calculated to be 1.6 nS. The mean estimated reversal potential for all basket-cell-mediated i.p.s.ps was -75.5 ± 7.6 mV ($n = 7$ cell pairs), and the mean conductance was estimated to be 1.5 ± 0.7 nS ($n = 6$), as described²⁸. The i.p.s.p. was blocked by bath application of 2 – 10 μ M bicuculline ($n = 2$). **B**, **a**, Firing of a pyramidal cell (lower trace), depolarized by constant current injection, becomes instantly entrained by periodic activation (4.7 Hz) of short trains of 3–4 action potentials in the presynaptic basket cell (upper trace, action potentials truncated). Dotted lines indicate intervals of 0.21 s. In all 5 basket and 2 axo-axonic cell to pyramidal cell connections tested for entrainment, the postsynaptic pyramidal cell became instantly phase-locked to the interneuron activity, firing at opposite phase to the interneuron. By entraining the postsynaptic pyramidal cell, the interneurons could decrease or increase the mean discharge rate of the pyramidal cell. Autocorrelation of the trace before (b) and during (c) regular basket cell activity shows an increased rhythmicity of pyramidal cell firing during periodic activation of the interneuron. **C**, **a**, Effect of periodic rhythmic discharge of a single basket cell on sub-threshold membrane potential oscillations in a pyramidal cell. Dotted lines indicate intervals of 0.26 s. Autocorrelograms of the trace before (b) and



intrinsic membrane conductances. Several voltage-dependent conductances have been reported to be involved in θ -frequency subthreshold membrane-potential oscillations in CA1 pyramidal cells¹³, as well as other cortical neurons^{18,19}. As these membrane-potential oscillations and the rebound depolarization occur in a similar voltage range, we investigated whether the entrainment observed here might be due to interaction with intrinsic oscillatory mechanisms. In pyramidal cells depolarized to fire action potentials, unitary perisomatic i.p.s.ps were repeatedly evoked. The synaptic response suppressed, and thereafter facilitated action-potential discharge in the pyramidal cell, as predicted⁶. Subsequent multiple clusters of action potentials following a single i.p.s.p. indicate that the hyperpolarizing response reset an intrinsic rhythmic state rather than simply delaying the onset of action-potential generation ($n = 10$ of 13 ; Fig. 2C). That this action of the i.p.s.ps was indeed due to resetting the phase of an intrinsic oscillatory state is apparent from the ability of i.p.s.ps to either reset or initiate subthreshold oscillations (Fig. 2D).

That single basket and axo-axonic cells can entrain pyramidal cell activity both at sub- and suprathreshold membrane potentials, suggests that synchronization of principal cell activity may be a fundamental role for these interneurons. We therefore tested, in the presence of blockers of excitatory amino-acid receptors, whether minimal stimulation of GABAergic interneurons could synchronize the firing of two spontaneously active pyramidal cells. Indeed, one single i.p.s.p. reset regular firing in both pyramidal cells ($n = 5$; Fig. 3A), whereas repeated minimal

following (c) entrainment show an increased rhythmicity during periodic activation of the interneuron.

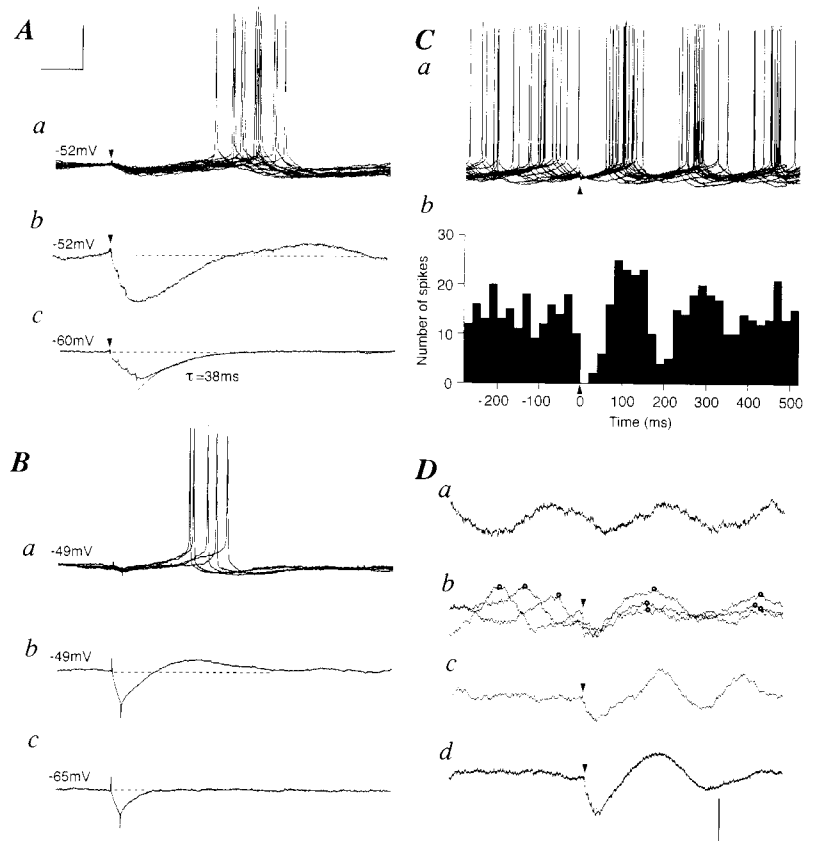
METHODS. Dual intracellular recordings were made from synaptically coupled CA1 interneuron to pyramidal cell pairs from 400 - μ m-thick transverse adult rat hippocampal slices as described⁸. Slices were maintained at the interface between artificial cerebrospinal fluid containing (in mM): 126 NaCl, 3 KCl, 1.25 NaH_2PO_4 , 24 NaHCO_3 , 2 MgSO_4 , 2 CaCl_2 , 10 glucose, pH 7.2 , and a humidified atmosphere of 95% O_2 and 5% CO_2 , at 34 – 35 $^\circ\text{C}$. Sharp glass microelectrodes contained 2% biocytin in 1.5 M potassium methylsulphate and were bevelled to a d.c. resistance of 60 – 150 M Ω . Recovered biocytin-labelled cells were analysed in the light microscope and target selectivity was determined by electron microscopy for a total of 94 basket cell and 22 axo-axonic cell synaptic contacts, ranging from 6 to 24 contacts per cell. All values are expressed as means $\pm$ s.d. The number of synaptically coupled pairs exceeds the number of interneurons because the latter could serve as the presynaptic partner for more than one recorded pyramidal cell.

stimulation at θ -frequencies phase-locked both neurons (Fig. 3B). Thus, GABAergic interneurons can effectively synchronize pyramidal cell activities. The exact contribution of different subtypes of interneuron remains to be established as they may be differentially active during different behavioural states. Anatomically identified basket cells have been reported to fire phase-related to extracellularly recorded θ oscillations¹⁶, whereas no such data are yet available for axo-axonic cells. As suggested by rhythmically occurring i.p.s.ps in basket cells during θ activity¹⁶, the basket cells themselves may be phased by GABAergic inputs from the medial septum²⁰ or from other interneurons in the hippocampal network. In addition, excitation through recurrent axon collaterals from the pyramidal cells^{8,21} would strengthen rhythmic basket cell discharges, suggesting a role for such a feedback loop²² in synchronization of cortical pyramidal cells²³.

Our results suggest that post-inhibitory 'rebound' activation of cortical principal cells must be considered as a fundamental facet in the functional repertoire of GABAergic interneurons acting on GABA_A receptors²⁴. The interaction of the synaptic events with intrinsic conductances provides a powerful mechanism, such that one single interneuron is sufficient to synchronize a large neuronal population. Given the ubiquitous nature of perisomatic GABAergic innervation throughout cortical structures²⁵, GABAergic phasing may represent a general mechanism for synchronization of cortical activity⁶. Indeed, intracellular recordings *in vivo* show that, at several distinct frequencies, rhythmic oscillatory activity is associated with periodic

FIG. 2 Mechanism underlying phasing of pyramidal cells.

A, Effect of a short train of 3–4 action potentials (position indicated by filled arrow) elicited in the presynaptic basket cell on a pyramidal cell at varying membrane potentials. Scale (vertical) represents 20 mV in panels *a* of **A**, **B** and **C** and 2 mV in panels *b*, *c* of **A** and **B**; horizontal, 50 ms in **A**, **B** and 100 ms in **C**. *a*, Superimposition of 12 consecutive sweeps in which the i.p.s.p. was followed by an action potential. *b*, Average of 23 sweeps showing subthreshold rebound depolarization following the i.p.s.p. Note that the action potentials shown in a cluster in a region temporally coinciding with the rise of the depolarizing overshoot. *c*, Postsynaptic response evoked at a more hyperpolarized membrane potential showing a monoexponential decay and no depolarizing overshoot. **B**, Effect of brief hyperpolarizing current pulses in a pyramidal cell. *a*, Superimposition of 5 single sweeps in which a small hyperpolarizing current pulse (0.05 nA) elicited a rebound action potential. *b*, Average of 333 sweeps showing subthreshold rebound depolarization following a short hyperpolarizing current pulse. *c*, Identical current pulses at more hyperpolarized membrane potentials failed to produce a depolarizing overshoot. **C**, Effect of unitary i.p.s.ps on the firing of the postsynaptic pyramidal cell. *a*, 19 consecutive sweeps aligned on the rising phase of single action potentials (trigger point indicated by filled triangle) evoked at 0.75 Hz in the presynaptic basket cell. *b*, Corresponding spike histogram for a total of 132 sweeps with 572 spikes. Note the regular occurrence of peaks in the histogram following the i.p.s.p. Bin width, 20 ms. **D**, Effect of unitary i.p.s.p. on subthreshold membrane potential oscillations of a postsynaptic pyramidal cell. *a*, Single sweep showing subthreshold membrane potential oscillations of the pyramidal cell. *b*, Three superimposed sweeps demonstrating the effect of single unitary i.p.s.ps (triangle) on subthreshold oscillations in the pyramidal cell. The positive peak of each cycle has been marked with an open circle. At whichever phase or membrane potential the i.p.s.p. occurred, the i.p.s.p. effectively reset the oscillation. *c*, Single sweep demonstrating that a unitary i.p.s.p. can initiate oscillations in the pyramidal cell when no obvious oscillatory activity was present before the i.p.s.p. *d*, Average of 15 sweeps without action potentials.



Subthreshold oscillatory activity occurring at random phases resulted in a flat pre-event baseline. Following the i.p.s.p., a characteristic waveform, consistent with the i.p.s.p. resetting an intrinsic oscillatory state. Scale in **D**: (vertical) *a*–*c*, 4 mV and *d*, 2 mV; horizontal is 100 ms for *a*–*d*.

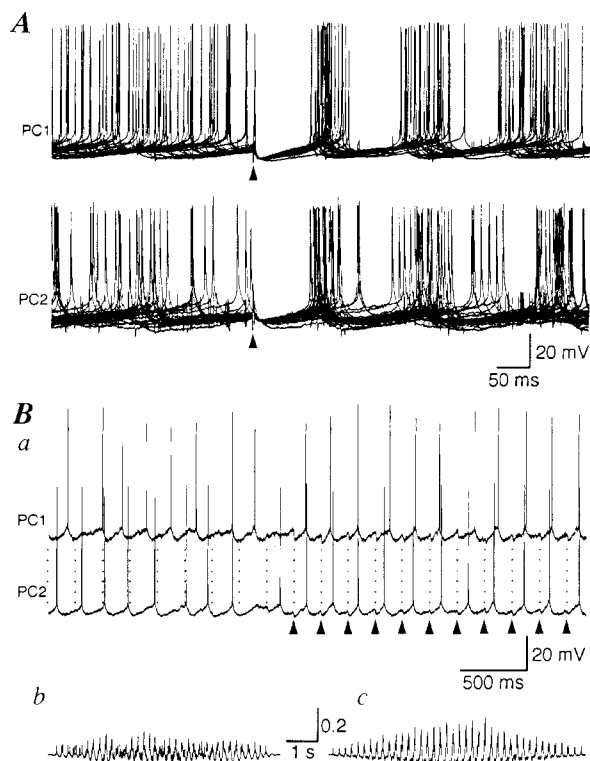


FIG. 3 Synchronization of pyramidal cell (PC) firing in the presence of ionotropic glutamate-receptor antagonists. **A**, Two simultaneously recorded pyramidal neurons were depolarized to elicit action potentials during which single i.p.s.ps (triangles), evoked at 0.2–0.5 Hz by minimal stimulation, reset the regular firing of both cells (30 consecutive sweeps; $n=5$). The stimulation strength was adjusted to evoke an i.p.s.p. of amplitude equivalent (<3 mV) to that produced by an individual, intracellularly recorded interneuron. In addition, rebound depolarization as in Fig. 2*A*, *b* could be evoked in both pyramidal cells (data not shown). **B**, *a*, Rhythmic i.p.s.ps evoked by minimal stimulation at 5 Hz (triangles) synchronize the firing of two simultaneously recorded pyramidal neurons. Dotted lines indicate intervals of 0.2 s. *b*, Cross-correlogram for the two neurons in a 5-s period before rhythmic minimal stimulation. *c*, Corresponding cross-correlogram for 5-s period following the start of rhythmic minimal stimulation. Note more pronounced cross-correlation during entrainment.

METHODS. Dual intracellular recordings were made from unconnected pairs of pyramidal cells separated by ~ 100 μ m. Minimal stimulation with a monopolar sharpened tungsten electrode (100 μ s, 3–8 V) was made in the pyramidal cell layer at a distance of ~ 100 μ m lateral from one of the recording electrodes in the presence of 10 μ M 6-cyano-7-nitroquinoxaline-2,3-dione and 30 μ M DL-2-amino-5-phosphonopentanoic acid.

i.p.s.ps, both in the hippocampus^{15,16,26} and the neocortex²⁷. We therefore suggest that a subset of cortical interneurons serve a key role in synchronizing principal cell activity, thus providing the temporal reference relative to which specific information may be coded. □

Received 30 May; accepted 7 September 1995.

1. Singer, W. A. *Rev. Physiol.* **55**, 349–374 (1993).
2. Mainen, Z. F. & Sejnowski, T. J. *Science* **268**, 1503–1506 (1995).
3. O'Keefe, J. & Recce, M. L. *Hippocampus* **3**, 317–330 (1993).
4. Lisman, J. E. & Idiart, M. A. P. *Science* **267**, 1512–1515 (1995).
5. Hopfield, J. J. *Nature* **376**, 33–36 (1995).
6. Lytton, W. W. & Sejnowski, T. J. *J. Neurophysiol.* **66**, 1059–1079 (1991).
7. Whittington, M. A., Traub, R. D. & Jefferys, J. G. R. *Nature* **373**, 612–615 (1995).
8. Buhl, E. H., Halasy, K. & Somogyi, P. *Nature* **368**, 823–828 (1994).
9. Li, X. G., Somogyi, P., Tepper, J. M. & Buzsáki, G. *Expl. Brain Res.* **90**, 519–525 (1992).
10. Sik, A., Penttonen, M., Ylinen, A. & Buzsáki, G. *J. Neurosci.* (in the press).
11. Niedermeyer, E. & Lopes da Silva, F. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields* (Williams & Wilkins, Baltimore, 1993).
12. Vanderwolf, C. H. *Electroencephalogr. clin. Neurophysiol.* **26**, 407–418 (1969).
13. Leung, L. W. & Yim, C. Y. *Brain Res.* **553**, 261–274 (1991).
14. Fox, S. E. *Expl. Brain Res.* **77**, 283–294 (1989).
15. Soltesz, I. & Deschenes, M. J. *Neurophysiol.* **70**, 97–116 (1993).
16. Ylinen, A. et al. *Hippocampus* **5**, 79–90 (1995).
17. Kandel, E. R. & Spencer, W. A. *J. Neurophysiol.* **24**, 243–259 (1961).
18. Alonso, A. & Llinas, R. R. *Nature* **342**, 175–177 (1989).
19. Llinas, R. R. *Science* **242**, 1654–1664 (1988).
20. Freund, T. F. & Antal, M. *Nature* **336**, 170–173 (1988).
21. Gulyás, A. I. et al. *Nature* **366**, 683–687 (1993).
22. Andersen, P., Eccles, J. C. & Loynning, Y. *Nature* **198**, 540–542 (1963).
23. Andersen, P. & Eccles, J. C. *Nature* **196**, 645–647 (1962).
24. Mody, I., De Koninck, Y., Otis, T. S. & Soltesz, I. *Trends Neurosci.* **17**, 517–525 (1994).
25. Somogyi, P. in *Neural Mechanisms of Visual Perception* (eds Lam, D. M.-K. & Gilbert, C. D.) 35–62 (Portfolio, The Woodlands, Texas, 1989).
26. Ylinen, A. et al. *J. Neurosci.* **15**, 30–46 (1995).
27. Steriade, M., Nunez, A. & Amzica, F. J. *Neurosci.* **13**, 3252–3265 (1993).
28. Ginsborg, B. L. *Biochim. biophys. Acta* **300**, 289–317 (1973).

ACKNOWLEDGEMENTS. We thank D. Latawiec and J. D. B. Roberts for technical assistance and A. D. Smith for his comments on the manuscript. K.H. was supported by the Wellcome Trust.

Mechanosensory signalling in *C. elegans* mediated by the GLR-1 glutamate receptor

Andres V. Maricq^{*†}, Erin Peckol^{*}, Monica Driscoll[‡] & Cornelia I. Bargmann^{*§}

^{*} Howard Hughes Medical Institute, Programs in Developmental Biology, Neuroscience and Genetics, Department of Anatomy, University of California, San Francisco, California 94143-0452, USA
[‡] Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08855, USA
[†] Present address: Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

NEURONAL signalling across synapses involves activation of many neurotransmitter receptors on postsynaptic cells. *glr-1* encodes a potential glutamate receptor in the nematode *Caenorhabditis elegans* which is most similar to vertebrate AMPA-type ionotropic glutamate receptors¹. *glr-1* is expressed in motor neurons and interneurons, including interneurons implicated in the control of locomotion². Here we investigate the contribution of *glr-1* to the normal signalling of these neurons, by generating a deletion mutation in *glr-1*. We find that mutant worms are deficient in their ability to withdraw backwards when mechanically stimulated, but they withdraw normally in response to chemical repellents. The ASH sensory neurons mediate withdrawal responses both to mechanical stimuli and to chemical repellents^{3,4}, and ASH makes chemical synapses with *glr-1*-expressing interneurons. Our results

suggest that postsynaptic interneurons use different neurotransmitter receptors to process two sensory stimuli detected by one sensory neuron.

The neurotransmitter glutamate is pivotal in synaptic transmission, synaptic plasticity and neuronal disease. The many types of glutamate receptors include the ionotropic glutamate receptors, ligand-gated ion channels that mediate rapid excitatory neurotransmission. In the rat, sixteen ionotropic receptor subtypes have been identified¹, not including variants of receptors introduced by alternative splicing and RNA editing, but the contribution of individual receptor subunits to neuronal signalling and behaviour is not well understood.

To study glutamate receptor function in *C. elegans*, we used the degenerate polymerase chain reaction (PCR) to identify a full-length complementary DNA for a potential *C. elegans* glutamate receptor, GLR-1. The predicted polypeptide sequence of 962 amino acids (Fig. 1a) shows the greatest identity to members of the AMPA α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid subclass of receptors^{5,6}, the *Drosophila* kainate-selective glutamate receptor⁷, and a putative receptor from *Lymnaea stagnalis*⁸ (Fig. 1b). The putative transmembrane regions are highly conserved between GLR-1 and AMPA receptors (Fig. 1c). The predicted ligand-binding domain of GLR-1, like vertebrate glutamate receptors, is similar to the *Escherichia coli* periplasmic glutamine-binding protein⁹. Based on this similarity, *glr-1* is likely to encode a receptor subunit for glutamate or a related ligand.

The expression pattern of *glr-1* was examined in transgenic worm strains that expressed reporter genes under the control of *glr-1* upstream sequences¹⁰ (Fig. 2a–c). Expression was limited to a subset of inter- and motor neurons, including the interneurons AVA, AVD, AVB and PVC. These 'command' interneurons are essential for normal locomotion and normal responses to mechanical stimuli². Polyclonal antisera raised against the GLR-1 protein recognized endogenous *C. elegans* antigen localized to neuronal processes (data not shown).

To define the function of the *glr-1*-expressing neurons, the toxic protein *mec-4(d)* was expressed under control of the *glr-1* promoter. This dominant allele of *mec-4* causes degenerative cell death of the sensory neurons in which it is normally expressed¹¹. MEC-4 is related to the vertebrate amiloride-sensitive sodium channel^{11,12}. In transgenic worms that ectopically expressed *mec-4(d)*, the *glr-1*-expressing neurons had a swollen, vacuolated appearance which is characteristic of degenerative death (Fig. 3a, b). These worms no longer responded to touch and had greatly impaired locomotion (Fig. 3c, d) confirming the importance of these neurons for coordinated movement.

A deletion mutation in *glr-1* was generated via insertion and imprecise excision of the transposon Tc1¹³. *glr-1(ky176)* encodes a predicted protein lacking 703 amino acids of the native receptor (Fig. 4a, b). *glr-1(ky176)* mutants exhibited defects in mechanosensory behaviour. They failed to respond to light touch to the nose (the Not phenotype¹⁶) (Fig. 4d). Locomotion of these animals was normal but slightly sluggish. The Not and sluggish phenotypes were rescued by injection of a clone of the wild-type *glr-1* gene into *glr-1(ky176)* mutants, indicating that they were caused by the *glr-1* mutation. We infer that *glr-1* is not required for all functions of the interneurons, because the *glr-1* mutant phenotype was much more subtle than the *glr-1::mec-4(d)* phenotype.

Other sensory modalities were not affected in *glr-1(ky176)* animals. Reversal in response to noxious chemicals (phenotype designated as Osm) was normal, demonstrating that *glr-1(ky176)* animals were not deficient in general reversal or locomotion. Spontaneous movement, chemotaxis, chemosensory adaptation, egg laying, dauer formation and defaecation also appeared normal in the mutants.

Worms respond to touch to the nose via the sensory neurons ASH (major input), and FLP and OLQ (minor input)³. All of these sensory neurons synapse directly onto *glr-1* expressing

§ To whom correspondence should be addressed.

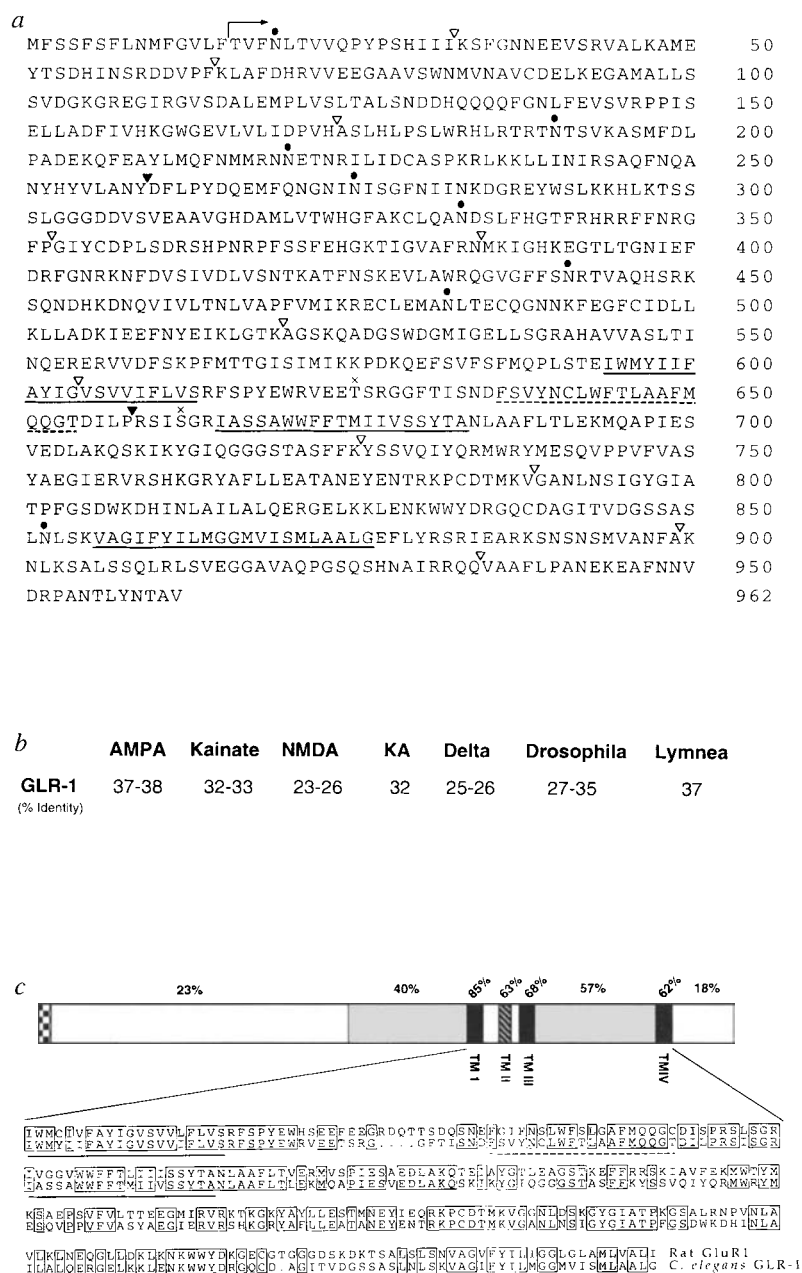


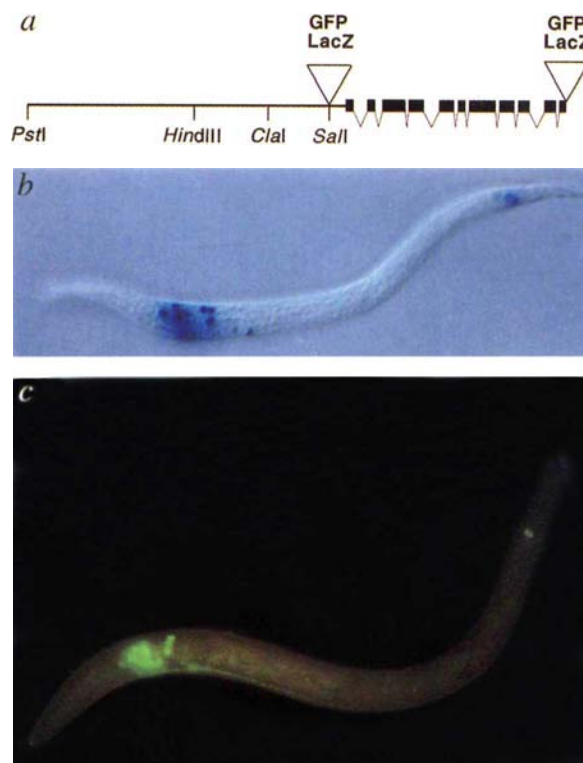
FIG. 1 a, Deduced amino-acid sequence of *glr-1* (single-letter code). Amino acids are numbered beginning with the first predicted methionine. An arrow indicates the predicted amino-terminus. Consensus N-linked glycosylation sites are indicated by bold dots, and consensus protein kinase C phosphorylation sites are indicated by crosses. The amino-acid sequences corresponding to four hydrophobic domains are underlined: these are the transmembrane domains TMI, TMIII and TMIV (solid underline) and TMII, which probably does not span the membrane²⁵ (dashed underline). Splice sites, determined by comparison of *glr-1* cDNA with the genomic sequence²⁶, are indicated by the open arrowheads. Analysis of genomic sequence and cDNA sequences of *glr-1* did not reveal alternative splicing near TMIV or RNA editing in TM2. The first exon and seventh intron of *glr-1* are different from those previously predicted from the genomic sequence²⁶. Solid arrowheads mark the boundaries of the genomic DNA deleted in the *glr-1* mutant (Fig. 4). The GenBank accession number for *glr-1* is U34661. b, Per cent amino-acid identity of *C. elegans* GLR-1 with members of other classes of glutamate receptors (GenBank accession numbers follow in parentheses). AMPA: rat GluR1–GluR4 (M36418–M36421); kainate: rat GluR5–GluR7 (Z11713, Z11715–Z11716); KA: rat KA-1, KA-2 (X59996, Z11581); Delta: rat delta-1, delta-2 (Z17238, Z17239); NMDA: rat

NMDAR1, rat NMDA2A–NMDA2D, *Drosophila* DNMDAR-1 (X63255, M91561–M91563, D13213, X71790); *Drosophila*: DGluRI–DGluRII (M73271–M97192); *Lymnea*: *L. stagnalis* (X60086). c, Domain organization and comparison of *C. elegans* GLR-1 with rat GluR1. Numbers indicate per cent amino-acid identity. Checked box indicates putative signal sequence; black boxes and solid underlines are putative transmembrane regions TMI, TMIII and TMIV; obliquely hatched box and dashed underline indicate TMII or putative pore-forming region; grey boxes represent the region of homology to the *E. coli* periplasmic glutamine-binding protein (30% amino-acid identity; SwisProt: GLNH-ecoli)⁹. Boxed sequences in c are identical in GLR-1 and GluR1.

METHODS. Degenerate oligonucleotide primers similar to vertebrate AMPA-selective glutamate receptors were used to amplify homologous DNA sequences from *C. elegans* first-strand cDNA. A ~500-bp PCR product was used to isolate cDNAs from a *C. elegans* cDNA library²⁷. The cDNA clones encoded nucleotides 40–3,003 of *glr-1*. The authentic 5' end was identified by reverse transcribed PCR using *glr-1* and splice leader SL1-specific oligonucleotides²⁷. Sequence analysis was assisted by the Genetics Computer Group, Wisconsin package²⁹ and the BLAST network service³⁰.

FIG. 2 *glr-1* is expressed in interneurons and motor neurons. **a**, Fusion genes used to study the expression of *glr-1*. Both transcriptional and translational fusions to β -galactosidase (LacZ) and green-fluorescent protein (GFP) were generated. **b**, Lateral view of a transgenic animal expressing LacZ (dark blue) under control of the *glr-1* promoter. **c**, Lateral view of a transgenic animal expressing an *unc-76*-GFP fusion protein under the *glr-1* promoter. Note neuronal processes in the ventral cord. Neurons were identified by comparing the localization of the LacZ or GFP signal to the position of neuronal nuclei viewed using Normarski optics. The classes of neurons that expressed GFP included: interneurons AVA, AVB, AVD, AVE, PVC, AIB and motor neurons RMD, RIM, SMD. AVG, PVQ and URY stained faintly. Anterior is at left and dorsal is at the top in all pictures.

METHODS. The *glr-1* cDNA was localized to the YAC Y72D7 and then to the cosmid C06E1, which is located on linkage group III near *dpy-19*. C06E1 was later sequenced by the *C. elegans*-genome sequence project²⁶. 5.3 kb upstream of *glr-1* were used to drive the expression of GFP::UNC-76 in the *C. elegans* expression backbone pPD21.28 (refs. 10, 31). Fusion proteins with this UNC-76 fragment are expressed in the processes of cells (L. Bloom, unpublished observations). In variants of these constructs, we replaced the *unc-76* sequence with a nuclear localization signals (NLS) and GFP with LacZ. Translational fusions were constructed by the in-frame ligation of an 8.8 kb *PstI*-*NsiI* fragment into either the LacZ expression vector pPD22.11 (ref. 31) or the corresponding GFP expression vector¹⁰. Transgenic worms were generated by microinjection and the arrays integrated by γ -irradiation with $\sim 6,000$ R delivered at 330 R min⁻¹. Cell identification was based on the characteristic morphology and position of cell nuclei¹⁴.



neurons¹⁴. The ASH sensory neuron makes chemical synapses onto the AVA, AVB and AVD interneurons¹⁴ (Fig. 4c). The defect in nose touch observed in *glr-1(ky176)* mutants suggests that these synapses are glutamatergic. *glr-1::mec-4(d)* GFP staining appeared normal in *glr-1(ky176)* worms, suggesting that *glr-1*-expressing neurons were present in normal number and position (data not shown).

glr-1(ky176) mutants also were defective in their response to light touch to the body (Mec phenotype) but not to a heavier prodding of the body (known as Tab phenotype, for touch abnormal¹⁵). A point mutation in *glr-1* shares the Osm⁺ Not⁻ phenotype with *glr-1(ky176)*, but is fully Mec⁺ (ref. 16). All *glr-1(ky176)* phenotypes were semi-dominant, but *glr-1* is not haploinsufficient, because *nDf20/+* heterozygous animals are Not⁺ and Mec⁺. Interestingly, the Mec response is believed to depend primarily on electrical and not chemical synapses,

so it is unlikely to use glutamate receptors directly for synaptic transmission². The Mec phenotype can only be partially rescued and might result from dominant interference of the truncated *glr-1(ky176)* protein with the development or function of receptors, gap junctions or other molecules at the synapse¹⁷.

The ASH neurons have a dual sensory function, mediating both the nose-touch response and the osmotic-avoidance response^{3,4} (Fig. 4c). Curiously, the osmotic-avoidance response is normal in *glr-1(ky176)* mutants (Fig. 4d and legend). The osmotic-avoidance response in *glr-1* mutants is still mediated by the ASH sensory neurons; killing ASH in *glr-1* mutants gave an osmotic-avoidance defect similar to that caused by killing ASH in wild-type animals. Thus 19/19 *glr-1(ky176)* animals remained in a 4M fructose osmotic ring for 20 min¹⁸, whereas only 3/12 *glr-1(ky176)* animals with killed ASH remained in the osmotic

FIG. 3 **a, b**, *glr-1* controlled expression of the degenerin *mec-4(d)* causes a necrotic neuronal death. **a**, Wild-type animals. **b**, *glr-1::mec-4(d)*: ectopic expression of MEC-4(d) is manifested by the grossly swollen appearance of the *glr-1*-expressing neurons (arrows). Below is shown the construct used to make the integrated transgenic strain of *C. elegans*. **c, d**, Worms that express *mec-4(d)* in *glr-1*-expressing neurons show greatly decreased spontaneous movement. **c**, Eight wild-type or **d**, seven *glr-1::mec-4(d)* transgenic worms were photographed 10 s (wild type) or 5 min (*glr-1::mec-4(d)*) after being placed on an agar plate.

METHODS. The *HindIII*-*KpnI* genomic fragment containing the entire *mec-4(d)* coding region¹¹ was ligated to the 5.3-kb *glr-1* promoter in the expression vector pPD49.26 (ref. 31). Generation of integrated and germline-transformed worm stains was as described for Fig. 2. Three integrated and germline-transformed worm stains with similar morphological and behavioural defects were isolated. Cells that did not express *glr-1* appeared to be intact from their appearance in Normarski microscopy.

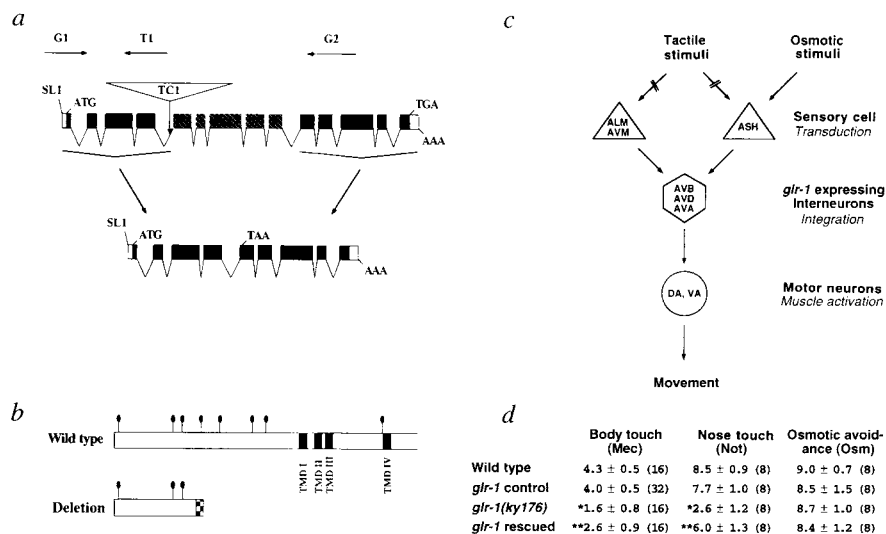
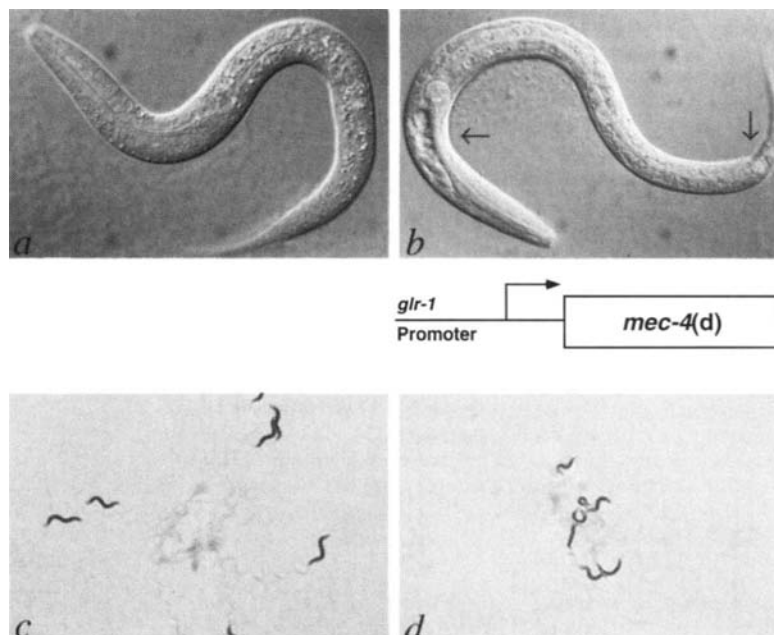


FIG. 4 Mechanosensory defects of the *glr-1* deletion mutation. **a**, Generation of the mutation. Exons are represented by filled boxes; the hatched region represents the part of the gene deleted during imprecise excision of the transposon Tc1. The deletion removes amino acids 260–659 and causes a frameshift at nucleotide 2,082 which then results in a premature stop codon at nucleotide 2,155 (Fig. 1a). SL1, splice leader sequence²⁸. **b**, Predicted protein product. *glr-1(ky176)* encodes a small N-terminal protein that lacks the putative glutamate-binding domains and transmembrane regions⁹, with a frameshifted region at the checkered box. Lollipops represent potential N-glycosylation sites. **c**, Inferred functions and connectivity associated with *glr-1*. Patterns of connectivity are simplified from ref. 14. **d**, Behavioural characterization of *glr-1* mutants. Wild-type animals, *glr-1* animals, and matched strain and rescued *glr-1* controls were tested for avoidance behaviours. Mechanosensation²: individual animals were scored for the number of reversals out of five trials (errors are s.e.m.; numbers of animals tested are given in parentheses); nose touch³: the average number of positive responses out of ten trials is given; osmotic avoidance¹⁸: the number of animals, out of ten per assay, that were confined to a 4M fructose ring after 15 min is indicated; total number of assays (10 animals each) is given in parentheses. Under these conditions, fewer than 20% of worms mutant for *osm-9(ky10)* remained in the ring. In addition, a quantitative osmotic-avoidance assay was produced by using a linear gradient of 0.5–0.35 M NaCl on a square agar plate, and the distribution of animals determined after a 30-min incubation on the plate; wild-type and *glr-1(ky176)* animals showed indistinguishable patterns of distribution in this linear gradient. Single asterisks denote responses that differ from wild type at $P < 0.01$ by a *t*-test (Statview II); double asterisks are responses of rescued *glr-1* animals that differ from the *glr-1* strain.

METHODS. A PCR-based screen was used to identify worms with a Tc1 transposon in the *glr-1* gene. MT3126 (*mut-2; dpy-19*) was screened for transposon Tc1 inserts in *glr-1*¹³. We used Tc1 specific primers and *glr-1* specific primers: G1a, GGAGACATCTTCGTACTCGAACC; G1b, CGATTGTGCATCTCCGAACGAC (in exon 12); G2a, TTCCTCGCTCAATCC-



TACTTCG; and G2b, CAACTAGCAGATGACCCGTCAAC (in exon 4). A Tc1 insertion into the fourth intron (TATATTA-Tc1-TAATTAA) imprecisely excised to produce a deletion that spanned introns 4–9 (intron 4)/... GTGCCAC– 1.62-kb deletion–CGTACTA.../intron 9). The *glr-1(ky176)* deletion was backcrossed 10 times. RT-PCR confirmed the presence of the predicted deleted mRNA. Germline rescue of *glr-1(ky176)* was obtained with a 10.7-kb genomic DNA fragment containing the *glr-1* gene. The *glr-1(ky176)* mutant strain includes a weak temperature-sensitive mutation in *dpy-19*; so the parental *glr-1* Tc1 insertion strain (Tc1 *ky175*) was used as a control. We cannot exclude a contribution of *dpy-19* to the phenotypes described here.

ring (χ^2 -20, $P < 0.001$). These results indicate that *glr-1* is required for the ASH neurons to initiate mechanical, but not chemical, avoidance behaviour. As *glr-1* is expressed in postsynaptic cells, mechanical and chemical stimuli must generate qualitatively or quantitatively distinct signals to the postsynaptic targets.

ASH neurons contain both clear and dark synaptic vesicles, indicating that they might release multiple transmitters¹⁴. One possibility is that osmotic signalling and mechanosensory signalling by ASH are mediated by distinct neurotransmitters. Most neurons in the vertebrate central nervous system contain several neurotransmitters—typically one classical neurotransmitter like glutamate and one or more peptide or catecholamine transmitters¹⁹. Different stimulation conditions preferentially

stimulate release of different transmitter types²⁰ which may result in either convergent²¹ or divergent effects on the postsynaptic cell²². Alternatively, functional glutamate receptors containing subunits other than GLR-1 might still be produced in *glr-1* mutants^{23,24}. In this model, different patterns of glutamate release by ASH in response to chemical or mechanical stimulation would activate different subtypes of glutamate receptors. These differences might also be interpreted in the context of other sensory neurons activated by chemical or mechanical stimuli. The discrete function of the GLR-1 receptor in behaviour reveals an unexpected complexity in *C. elegans* synaptic signalling; we speculate that GLR-1 contributes to sensory discrimination between mechanical and chemical stimuli by the animal. □

Received 12 May; accepted 28 August 1995.

- Hollmann, M. & Heinemann, S. *Rev. Neurosci.* **17**, 31–108 (1994).
- Chalfie, M. et al. *J. Neurosci.* **5**, 956–964 (1985).
- Kaplan, J. & Horvitz, H. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2227–2231 (1993).
- Bargmann, C., Thomas, J. & Horvitz, H. *Cold Spring Harb. Symp. quant. Biol.* **55**, 529–538 (1990).
- Keinänen, K. et al. *Science* **249**, 556–560 (1990).
- Boulter, J. et al. *Science* **249**, 1033–1037 (1990).
- Ullrich, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10484–10488 (1992).
- Darlington, M. G., Hutton, M. L. & Harvey, R. J. in *Exs 63: Comparative Molecular Neurobiology* (ed. Pichon, Y.) 48–64 (Birkhäuser, Basel, 1993).
- Stern-Bach, Y. et al. *Neuron* **13**, 1345–1357 (1994).
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. & Prasher, D. *Science* **263**, 802–805 (1994).
- Driscoll, M. & Chalfie, M. *Nature* **349**, 588–593 (1991).
- Canessa, C. M., Harrisberger, J. D. & Rossier, B. C. *Nature* **361**, 467–470 (1993).
- Zwaal, R. R., Broeks, A., van Meurs, J., Groenen, J. T. & Plasterk, R. H. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7431–7435 (1993).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. *Phil. Trans. Soc. Lond. B* **314**, 1–340 (1986).
- Chalfie, M. & Wolinsky, E. *Nature* **345**, 410–416 (1990).
- Hart, A. C., Sims, S., Kaplan, J. M. *Nature* **378**, 82–85 (1995).
- Verrall, S. & Hall, Z. W. *Cell* **68**, 23–31 (1992).
- Culotti, J. G. & Russell, R. L. *Genetics* **90**, 243–256 (1978).

- O'Malley, D. M., Sandell, J. H. & Masland, R. H. *J. Neurosci.* **12**, 1394–1408 (1992).
- DeVries, S. H. & Baylor, D. A. *Cell/Neuron* **72/10** (suppl.), 139–149 (1993).
- Schotland, J. et al. *Nature* **374**, 266–268 (1995).
- Church, P. J., Whim, M. D. & Lloyd, P. E. *J. Neurosci.* **13**, 2790–800 (1993).
- Bochet, P. et al. *Neuron* **12**, 383–388 (1994).
- Jonas, P., Racca, C., Sakmann, B., Seeburg, P. H. & Monyer, H. *Neuron* **12**, 1281–1289 (1994).
- Hollmann, M., Maron, C. & Heinemann, S. *Neuron* **13**, 1331–1343 (1994).
- Wilson, R. et al. *Nature* **368**, 32–38 (1994).
- Barstead, R. J. & Waterston, R. H. *J. Biol. Chem.* **264**, 10177–10185 (1989).
- Krause, M. & Hirsh, D. I. *Cell* **49**, 753–761 (1987).
- Devereaux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
- Altshul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. molec. Biol.* **215**, 403–410 (1990).
- Fire, A., White-Harrison, S. & Dixon, D. *Gene* **93**, 189–198 (1990).

ACKNOWLEDGEMENTS. We thank A. Hart and J. Kaplan for communicating results before publication and for their assistance in scoring the Not phenotype; S. Salser and J. Maloof for advice on antibody staining; R. Barstead for the cDNA library; H. Colbert for adaptation assays; and L. Avery and members of C.I.B.'s laboratory for comments on the manuscript. This work was supported by a NIH grant to C.I.B. A.V.M. is the recipient of an ADAMHA Scientist Development Award; C.I.B. is a Lucille P. Markey Scholar, a Searle Scholar, and an assistant investigator of HHMI.

Synaptic code for sensory modalities revealed by *C. elegans* GLR-1 glutamate receptor

Anne C. Hart, Shannon Sims & Joshua M. Kaplan*

Department of Molecular Biology, Massachusetts General Hospital, 50 Blossom Street, Boston, Massachusetts 02114, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

How does the nervous system encode environmental stimuli as sensory experiences? Both the type (visual, olfactory, gustatory, mechanical or auditory) and the quality of a stimulus (spatial position, intensity or frequency) are represented as a neural code. Here we undertake a genetic analysis of sensory modality coding in *Caenorhabditis elegans*. The ASH sensory neurons respond to two distinct sensory stimuli (nose touch and osmotic stimuli). A mutation in the *glr-1* (glutamate receptor) gene eliminates the response to nose touch but not to osmotic repellents. The predicted GLR-1 protein is roughly 40% identical to mammalian AMPA-class glutamate receptor (GluR) subunits. Analysis of *glr-1* expression and genetic mosaics indicates that GLR-1 receptors act in synaptic targets of the ASH neurons. We propose that discrimination between the ASH sensory modalities arises from differential release of ASH neurotransmitters in response to different stimuli.

We studied a simple neural circuit that encodes the response to two sensory stimuli, both of which elicit backward locomotion (Fig. 1a). To identify genes required to distinguish the ASH modalities, we screened for mutations that disrupt the ASH-mediated touch response but that have no effect on responsiveness to osmotic repellents (J.M.K. and H. R. Horvitz, unpublished observations). We isolated the *n2461* mutation which defined the *glr-1* gene (Table 1a). *glr-1* homozygotes and *glr-1/nDf20* heterozygotes have identical phenotypes, which suggests that the *n2461* mutation causes a severe or complete loss of gene function. Because *glr-1* mutants respond normally to osmotic repellents, it is likely that the ASH neurons of mutant *glr-1* animals function normally in osmotic avoidance, but other sensory neurons may account for the osmotic sensitivity of *glr-1* mutants. To distinguish between these possibilities, we showed that killing the ASH neurons severely diminishes the response of *glr-1* mutants to an osmotic stimulus (Table 1a). We conclude that *glr-1* mutants are defective for ASH-mediated nose-touch sensitivity but are normal for ASH-mediated osmotic sensitivity. *glr-1* mutants are normal for many other behaviours, including chemotaxis, male mating, and dauer formation¹⁻³. The *glr-1* gene appears to play a relatively specific role in ASH-mediated touch sensitivity; however, further analysis revealed defects in two additional mechanosensory behaviours (described below).

Because it may play a role in modality coding, we further characterized the *glr-1* gene (Fig. 2). The *glr-1(n2461)* mutation maps between *unc-116* and *dpy-19* on chromosome III. The cosmid C06E1 and subcloned derivatives thereof (CX #1 and KP #1) rescued the touch sensitivity of *glr-1(n2461)* animals. The sequence of the minimal rescuing fragment (KP #1), determined by the *C. elegans* genome sequencing consortium⁴, contains a gene (*C06E1.4*; accession number L16559) that is highly similar to mammalian AMPA-class GluR genes (for example, 40% identical to rat GluRB). A frameshift mutation in this GluR gene (KP #33) eliminated the *glr-1* rescuing activity. The *n2461* mutation corresponds to a nonsense mutation in codon 807 and the predicted *n2461* GLR-1 protein lacks the 'flip/flop' domain⁵,

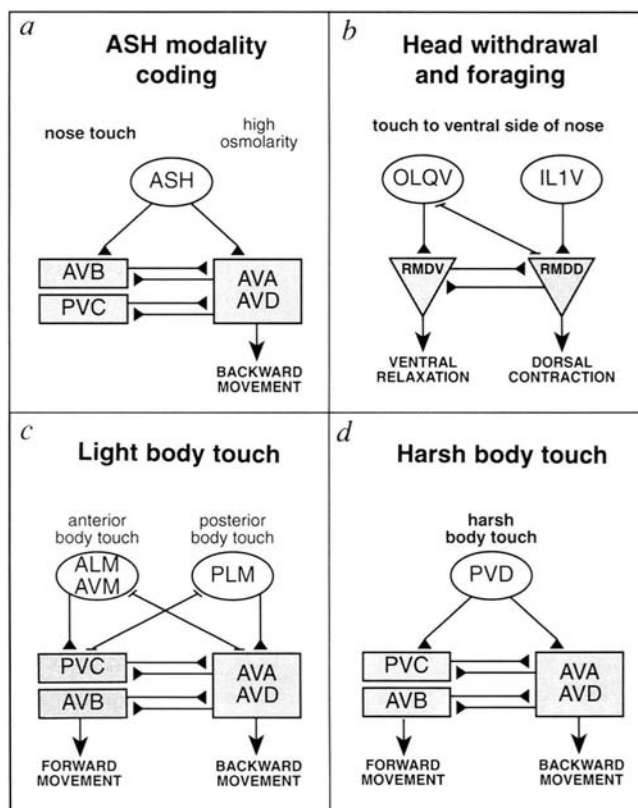


FIG. 1 Role of GLR-1 receptors in several behaviours. Sensory neurons are represented as ovals, interneurons as boxes and motor neurons as triangles. Cells that express *glr-1* are shown in grey. Synaptic connections are indicated by lines ending in triangles; gap junctions by lines ending in bars. Connectivities have been described⁹. To simplify the illustration, gap junctions between ASH and AVD and between PVC and AVA are not shown. Behaviours that are defective in *glr-1* mutants are shown in bold. a, *glr-1* mutants are defective for the nose-touch response but are normal for osmotic avoidance. The ASH neurons respond to two stimuli, both of which stimulate backward locomotion^{14,15}. The command neurons AVA and AVD drive backward movement whereas AVB and PVC drive forward movement⁷. b, *glr-1* is required for normal head withdrawal and foraging behaviours. The mechanosensory neurons IL1 and OLQ, and the RMD motor neurons regulate the rate of spontaneous foraging and mediate an aversive response to touch to the ventral or dorsal tip of the nose (J.K. and H.R. Horvitz, unpublished results). *glr-1* mutants forage abnormally slowly and are defective for the head-withdrawal reflex. c, *glr-1* animals respond normally to light touch to the body. The mechanosensory neurons ALM, AVM and PLM sense touch to the body and provide input to the command neurons via both synaptic connections and gap junctions⁷. Anterior and posterior body touch stimulate backward and forward locomotion, respectively. d, *glr-1* is required for harsh-touch sensitivity. Harsh touch to the body is mediated by the PVD sensory neurons, which provide input to AVA and PVC solely via synaptic connections^{8,9}.

the last transmembrane domain and the cytoplasmic tail (Fig. 2b). Do GLR-1 receptors act as GluRs? Treatment with AMPA, a glutamate agonist, caused hyperactive foraging behaviour (described below) in wild-type but not in *glr-1* animals (data not shown). We conclude that *glr-1* encodes a GluR and that the *n2461* mutation results in a severe loss of GLR-1 receptor function.

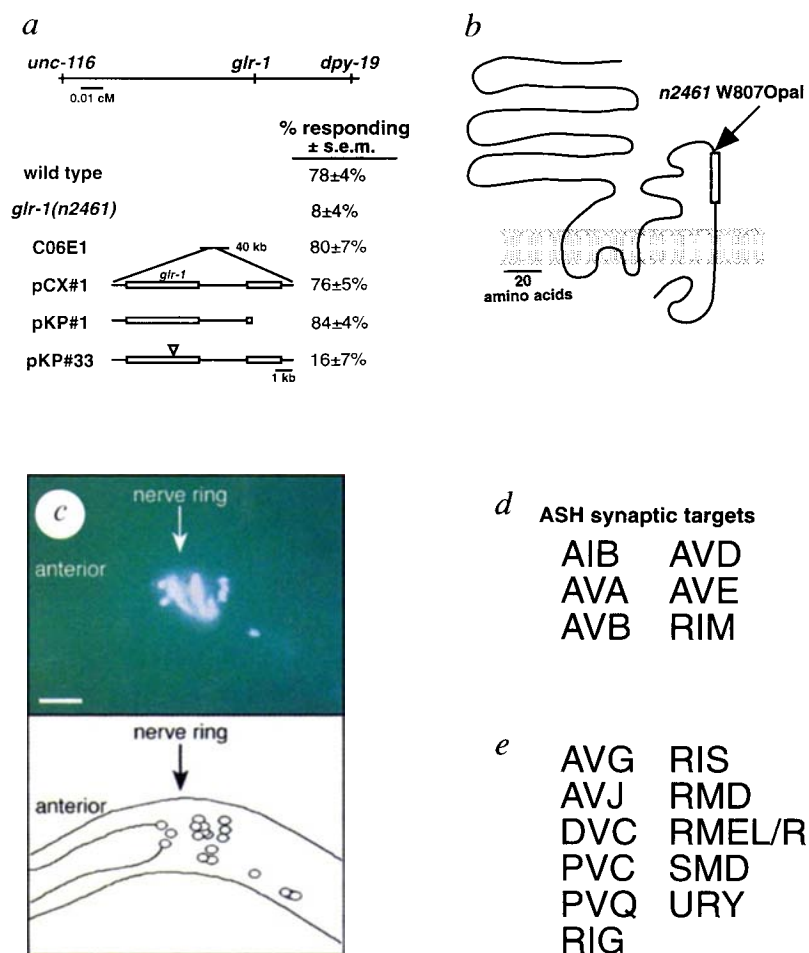
We determined the cellular expression pattern of the *glr-1* gene (Fig. 2c). A chimaeric *glr-1::gfp* gene containing 5' transcription control sequences of *glr-1* fused to the coding sequences of the green fluorescent protein (*gfp*) gene⁶ was expressed in transgenic animals. Seventeen classes of neurons, including several synaptic targets of ASH, expressed the *glr-1::gfp* gene (Fig. 2c, d); however, the ASH neurons did not express the *glr-1::gfp* gene. These results suggest that GLR-1 receptors act in synaptic targets of ASH in the nose-touch response (Fig. 1a, and see below). The expression pattern also suggests that GLR-1 receptors may be required for several other behaviours (Table 1b and Fig. 1b-d).

Because *glr-1* is expressed in the RMD motor neurons, which innervate head muscles, we examined *glr-1* mutants for changes

* To whom correspondence should be addressed.

FIG. 2 *glr-1* encodes an AMPA-type GluR expressed in 17 classes of neurons. **a**, *glr-1* maps between *unc-116* and *dpy-19*. The cosmid C06E1, and subclone derivatives of C06E1 (CX#1 and KP#1), rescued the nose touch response of *glr-1* mutants. A frameshift mutant (KP#33) lacked rescuing activity. **b**, The predicted topology of the GLR-1 GluR is illustrated (based on that determined for mammalian GluRs¹⁷). The n2461 mutation corresponds to a nonsense mutation of codon 807, which truncates the protein before the flip-flop homology (box) and the last transmembrane domain. **c**, Expression of the *glr-1::gfp* reporter construct (KP#6) in an adult animal visualized by fluorescence microscopy. Scale bar, 50 μ m. **d** and **e**, Identities of the *glr-1* expressing neurons are indicated. Neurons that receive synaptic input from ASH listed in **D**⁹.

METHODS. The n2461 mutation was genetically mapped as follows. Deficiency and duplication mapping: *glr-1* fails to complement *nDf20*; *sDp3* does not cover *glr-1*. Three-factor mapping: *unc-86(n846)* (9/18) *glr-1(n2461)* (3/18) *dpy-19(e1259ts)* (1/18) *sup-5(e1464)* (5/18) *unc-32(e189)*; *unc-116(e2281)* (3/6) *glr-1(n2461)* (2/6) *dpy-19(e1259ts)* (1/6) *unc-32(e189)* and one double recombinant of the genotype *unc-116(e2281) dpy-19(e1259ts) unc-32(e189)*; and *unc-116(e2281)* (2/5) *glr-1(n2461)* (3/5) *unc-32(e189)*. Cosmids corresponding to the genetic interval between *unc-116* and *dpy-19* were tested for rescue of the *glr-1* phenotype in transgenic animals. Transgenic animals were prepared by microinjection with either cosmids (2–10 ng μ l⁻¹) or plasmids (10–50 ng μ l⁻¹) using *lin-15* as a transformation marker¹⁸. Plasmid subclones were constructed as follows: CX#1 is a subclone of the 10 kb *XhoI/PstI* fragment from C06E1 in pBSII; KP#1 is a 8.5-kb subclone of CX#1 derived by deleting the *Bsu36I/BamHI* fragment; KP#33 was derived from CX#1 by filling in the *NcoI* site (at nucleotide 10,801 of C06E1) and re-ligation. The n2461 mutation was sequenced as follows: sequences from 500 nucleotides upstream of the initiator ATG through to the last predicted codon of *glr-1* were amplified from *glr-1* (n2461) mutants and subcloned in pBSII. Several subclones were sequenced. The n2461 mutation corresponds to a nonsense mutation in codon 807 and eliminates a *BsrI* restriction site. The chimeric *glr-1::gfp* gene (KP#6) was constructed by subcloning the 5.3-kb *PstI/SalI* fragment of CX#1 into the *gfp* plasmid TU#61⁶. Transgenic animals (KP537 *nuls1*) carrying KP#6 as an integrated transgene were analysed by fluorescence microscopy for the expression of *glr-1*. Expressing neurons were identified based on nuclear positions and



in head movements. The RMD neurons are required for two behaviours, head withdrawal and foraging (Fig. 1b). *C. elegans* responds to dorsal or ventral touch to the nose with an aversive head-withdrawal reflex. *glr-1* mutants were defective in the head-withdrawal reflex and this defect was rescued by the CX#1 plasmid (Table 1b). Wild-type animals move the tip of their nose in a rhythmic motion, termed foraging behaviour. *glr-1* mutants foraged slowly, whereas transgenic animals carrying extra copies of the wild-type *glr-1* gene foraged hyperactively (data not shown). Similarly, treatment with AMPA caused hyperactive foraging. These results demonstrate that GLR-1 receptors regulate the RMD neurons.

Interneurons required for forward (AVB and PVC) and backward (AVA and AVD) locomotion express *glr-1*, which suggests that GLR-1 receptors may play a role in regulating locomotion. These interneurons have been designated command neurons because they play a critical role in locomotion⁷. *glr-1* mutants are normal for spontaneous locomotion, which suggests that GLR-1 receptors are not required to maintain the basal activity of the command neurons. Because they are defective for the nose-touch response, we investigated whether *glr-1* mutants might be defective for locomotion elicited by other sensory stimuli. Light touch to the body is sensed by the mechanosensory

axonal morphologies^{9,10}. Neuronal class designations are as described⁹. These cellular identifications were also consistent with the patterns of *glr-1::gfp* expression observed in mosaic animals (Fig. 3). All cells listed expressed *glr-1::gfp* in L1 larvae. Some cells expressed *glr-1::gfp* only in later larval stages (for example, DVC, PVQ and AVJ). The DVB neurons also occasionally expressed *glr-1::gfp*. Expression was first apparent in threefold embryos.

neurons ALM, AVM, and PLM, which provide input to the command neurons through both synaptic connections and gap junctions (Fig. 1c)⁷. In contrast, harsh-touch sensitivity is mediated by the mechanosensory neuron PVD, which provides input to the command neurons via synaptic connections (Fig. 1d)^{8,9}. We found that *glr-1* mutants were normal for light-touch sensitivity but were defective for harsh-touch sensitivity (Table 1b). These results suggest that GLR-1 receptors may act at PVD-to-interneuron synapses. As light-touch sensitivity is likely to be mediated by gap junctions⁷, it is not surprising that *glr-1* mutants have normal light-touch sensitivity.

Where do GLR-1 receptors act in the nose-touch response? Because the ASH neurons do not appear to express *glr-1*, it seems likely that GLR-1 receptors act in synaptic targets of ASH. As nose touch elicits backward movement, the backward command neurons (AVA and AVD) are likely to be critical synaptic targets. However, ASH also provides synaptic input to several other classes of *glr-1*-expressing neurons (Fig. 2d), including the forward command neuron AVB. Furthermore, our expression studies do not exclude function of *glr-1* in the ASH neurons because *glr-1::gfp* expression may not accurately reflect the expression of the endogenous gene. To determine where GLR-1 receptors act in this circuit, we analysed the nose-touch

response of genetic mosaics (Fig. 3). The backward command neurons descend from the embryonic blastomere AB.a, whereas the forward command neurons and the ASH neurons descend from AB.p¹⁰. Loss of *glr-1* function in either AB.a or AB.p caused a partial defect, but total losses caused a severe defect in the nose-touch response. We further mapped the focus of *glr-1* action in AB.a to AB.al, which gives rise to both AVA and AVD. We conclude that GLR-1 receptors do not act solely in the backward command neurons or in the ASH neurons, in agreement with the expression studies.

Because the only *glr-1*-expressing cells that are known to play a role in locomotion are the command neurons, we propose that a normal nose-touch response requires functional GLR-1 receptors in both the forward and backward command neurons, thereby explaining the requirement in both AB.a and AB.p lineages. Nose-touch avoidance may require the coordinated activities of both the forward and backward locomotion circuits. Consistent with this idea, the forward and backward command neurons have multiple reciprocal synaptic connections (Fig. 1), which might coordinate the activities of these antagonistic circuits. As AMPA class GluRs (and hence GLR-1 GluRs) are excitatory, the ASH inputs to both forward and backward locomotion are probably excitatory. Previous laser microsurgery experiments suggested that AVB and PVC drive forward movement⁷; however, our analysis of *glr-1* suggests that AVB (and perhaps PVC) also plays a role in backward movement. As AVB and PVC have multiple synaptic partners, it is not surprising that eliminating a subset of their inputs causes a different defect from killing these cells altogether.

What is the mechanism of ASH modality coding? *glr-1* mutants have a selective defect in ASH-mediated touch sensitivity, yet GLR-1 receptors appear to act in the synaptic targets of the ASH neurons. We propose that the signals produced at the ASH-to-interneuron synapses following touch differ from those elic-

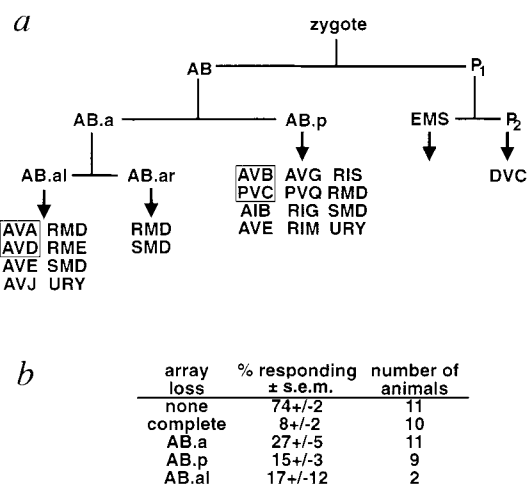


FIG. 3 GLR-1 receptors act in synaptic targets of ASH. Animals lacking *glr-1* function in either AB.a or AB.p lineages and non-mosaic controls were analysed for the nose-touch response (errors indicate the s.e.m.). *glr-1*-expressing cells derived from AB.al, AB.ar and AB.p are illustrated. Command interneurons are outlined by boxes (see text).

METHODS. Two different genetic backgrounds were used for mosaic analysis: KP820 *nc1-1(e1865) glr-1(n2461); lin-15(n765ts); nuEx146* and KP821 *nc1-1(e1865) glr-1(n2461); nuEx147*. Both strains carry transgenes containing *glr-1(+)*, two cell autonomous markers (*nc1-1(+)* and *glr-1::gfp*), and one of two transformation markers (*lin-15(+)* or KP#16, which confers a dominant Dpy phenotype). Because no statistical differences were observed between them, results from the two strains were combined. Mosaic and non-mosaic L3 and L4 larvae were isolated based on their *Ncl* phenotypes and their nose-touch responses were examined as young adults. Following the behavioural analysis, the patterns of mosaicism were confirmed by examining the *Ncl* phenotype and the expression of *glr-1::gfp*.

ited by osmotic repellents and that this differential output constitutes a synaptic code for sensory modalities. For example, in response to touch, the ASH neurons may release glutamate whereas, following osmotic stimuli, ASH neurons might release both glutamate and a peptide cotransmitter. Consistent with this hypothesis, ASH synaptic terminals contain both clear and dense core synaptic vesicles⁹, suggesting that ASH neurons use at least two distinct neurotransmitters. In several cases, differential release of neurotransmitters appears to result from differences in the magnitude or pattern of neuronal activity^{11,13}. This could provide a mechanism for producing a synaptic code for ASH sensory modalities. For example, if nose touch is a weaker stimulus than osmotic shock, then nose touch might elicit the release of only the clear core vesicles (which are closest to the active zone). By contrast, high osmolarity could stimulate release of both the clear and dense core vesicles.

Our results do not exclude other models for modality coding. If sensory stimuli produced a graded release of glutamate, ASH modalities could be distinguished by the amount of released glutamate. Alternatively, the sensory stimuli might activate different sets of sensory neurons. In this case, the distinct populations of responding sensory neurons could constitute a modality code. Further analysis of modality specific genes should allow us to elucidate the molecular basis of modality coding in this simple neural circuit. □

Received 12 May; accepted 18 August 1995.

1. Bargmann, C. *Cell* **74**, 515–527 (1993).
2. Hodgkin, J. *Genetics* **103**, 43–64 (1983).
3. Riddle, D. L. in *The Nematode C. elegans* (ed. Wood, W. B.) 393–412 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1988).
4. Sulston, J. et al. *Nature* **356**, 37–41 (1992).
5. Sommer, B. et al. *Science* **249**, 1580–1585 (1990).
6. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. & Prasher, D. *Science* **263**, 802–805 (1994).
7. Chalfie, M. et al. *J. Neurosci.* **5**, 956–964 (1985).
8. Way, J. C. & Chalfie, M. *Genes Dev.* **3**, 1823–1833 (1989).
9. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. *Phil. Trans. R. Soc. Lond.* **314**, 1–340 (1986).

TABLE 1 Behavioural analysis

(a) Role of <i>glr-1</i> in ASH sensory modalities					
Genotype	Cells killed	Nose touch (% responding)	Osmotic avoidance escape time (min)		
Wild type	None	86 ± 1 (10,100)	9.5 ± 0.8 (10,40)		
<i>glr-1(n2461)</i>	None	6 ± 3 (10,100)	8.3 ± 0.6 (10,40)		
<i>glr-1(n2461)</i>	ASH	ND	3.1 ± 0.4 (11,44)		
(b) Role of <i>glr-1</i> in other behaviours					
Genotype	Head withdrawal (% responding)	Foraging	Locomotion	Light body touch	Harsh body touch
Wild type	85 ± 4% (10,100)	+	+	+	+
<i>glr-1(n2461)</i>	2 ± 1% (10,100)	slow	+	+	–

The following behaviours were assayed as previously described: nose touch¹⁴, light body touch⁷, harsh body touch⁸, osmotic avoidance¹⁵. Errors indicate the s.e.m. and the numbers of animals and trials are shown in parentheses. All behavioural studies were done by an experimenter unaware of the animal's genotype and operative status. a, Nose touch was assayed as follows: animals were tested 10 times each, with a positive response being scored when animals either stop forward movement or initiate backward movement following the stimulus. The ASH neurons of L1 larvae were killed with a laser microbeam as described¹⁴. Forty-eight hours after recovery from operations, the osmotic avoidance behaviour of the animals was assayed by placing animals in 1-cm rings formed from 4M fructose and examining them once a minute to determine when they escaped the ring. For laser-operated animals, individual animals were placed in the ring and each animal was assayed 4 times. For unoperated controls, 10 wild-type or mutant *glr-1* animals were placed in each ring, and each group of animals was assayed 4 times. In addition, five unoperated mutant *glr-1* animals were scored individually (as laser-operated animals)—their escape times (8.8 $\pm$ 0.2 min) were not statistically different from the grouped controls. After assaying osmotic avoidance, laser killing of ASH neurons was confirmed by staining animals with the fluorescent dye DiO¹⁶. b, Head withdrawal was assayed by allowing the dorsal or ventral tip of the nose to touch an eyelash; a response was considered positive when the animal retracted its head from the stimulus. Harsh-touch sensitivity can be assayed only in animals that lack functional light-touch sensory neurons (such as *mec-7*). *glr-1(n2461);mec-7(u443)* double mutants were significantly less responsive to harsh touch than were control *mec-7* mutants.

10. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. *Devl Biol.* **100**, 64–119 (1983).
11. Peng, Y. & Zucker, R. S. *Neuron* **10**, 465–473 (1993).
12. DeVries, S. H. & Baylor, D. A. *Cell/Neuron* **72/10**, 139–149 (1993).
13. Whim, M. D. & Lloyd, P. E. *J. Neurosci.* **14**, 4244–4251 (1994).
14. Kaplan, J. M. & Horvitz, H. R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2227–2231 (1993).
15. Bargmann, C. I., Thomas, J. H. & Horvitz, H. R. *Cold Spring Harbor Symp. quant. Biol.* **55**, 529–538 (1990).
16. Herman, R. & Hedgecock, E. *Nature* **348**, 169–171 (1990).
17. Hollmann, M., Maron, C. & Heinemann, S. *Neuron* **13**, 1331–1343 (1994).
18. Huang, L., Tzou, P. & Sternberg, P. *Molec. Biol. Cell* **5**, 395–412 (1994).

ACKNOWLEDGEMENTS. We thank A. Coulson for cosmids, V. Maricq and C. Bargmann for plasmids and for sharing unpublished results, M. Chalfie for *gfp* plasmids, and B. Sawin and the members of the Ruvkun and J.M.K. labs for comments on the manuscript. This work was supported by grants to J.K. from the NIH and the Human Frontiers Science Program. A.C.H. was supported by a postdoctoral fellowship from the Muscular Dystrophy Association. J.K. is a Pew Scholar in the biomedical sciences.

Clustering of Shaker-type K⁺ channels by interaction with a family of membrane-associated guanylate kinases

Eunjoon Kim, Martin Niethammer,
Adam Rothschild, Yuh Nung Jan*
& Morgan Sheng[†]

Howard Hughes Medical Institute, Massachusetts General Hospital,
Department of Neurobiology, Harvard Medical School, Boston,
Massachusetts 02114, USA

* Howard Hughes Medical Institute, and

Departments of Physiology and Biochemistry, University of California,
San Francisco, California 94143, USA

ANCHORING of ion channels at specific subcellular sites is critical for neuronal signalling, but the mechanisms underlying channel localization and clustering are largely unknown (reviewed in ref. 1). Voltage-gated K⁺ channels are concentrated in various neuronal domains, including presynaptic terminals, nodes of Ranvier and dendrites, where they regulate local membrane excitability. Here we present functional and biochemical evidence that cell-surface clustering of Shaker-subfamily K⁺ channels is mediated by the PSD-95 family of membrane-associated putative guanylate kinases, as a result of direct binding of the carboxy-terminal cytoplasmic tails of the K⁺ channel subunits to two PDZ (also known as GLGF or DHR) domains in the PSD-95 protein². The ability of PDZ domains to function as independent modules for protein-protein interaction, and their presence in other junction-associated molecules (such as ZO-1 (ref. 3) and syntrophin⁴), suggest that PDZ-domain-containing polypeptides may be widely involved in the organization of proteins at sites of membrane specialization.

The amino-terminal intracellular region of voltage-gated K⁺ channel α -subunits contains a proximal domain involved in subunit multimerization⁵, and a distal flexible 'ball peptide' responsible for rapid inactivation of the channel⁶. We reasoned therefore that potential anchoring molecules would probably bind to the C-terminal intracellular tail of these K⁺ channel proteins. A yeast two-hybrid screen of a human brain complementary DNA library, using as bait the C-terminal tail of Kv1.4 (a member of the Shaker subfamily; ref. 7), yielded 5 distinct cDNAs encoding polypeptides that interacted specifically with the C-terminal tail of Kv1.4 (Fig. 1a). Clones 1, 2 and 3 were different overlapping cDNA fragments of PSD-95 (also known as SAP90)^{2,8}; clone 4 was a cDNA fragment of hdlg/SAP97 (a close relative of PSD-95)^{9,10}; and clone 5 was a fragment of a related but unknown gene with ~80% amino-acid identity to both PSD-95 and hdlg/SAP97 over its length (E.K., unpublished

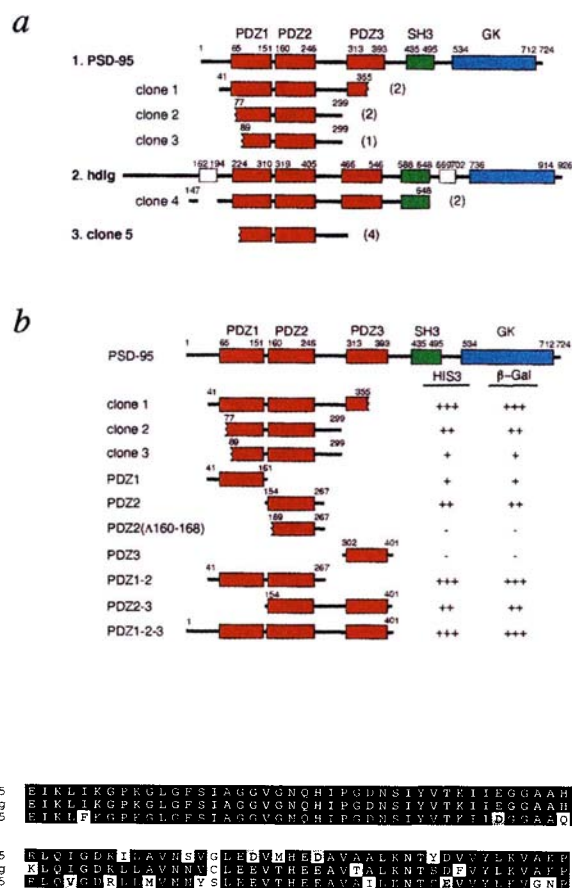


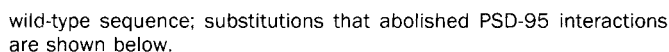
FIG. 1 a, The five Kv1.4C-interacting clones isolated by the yeast two-hybrid screen, showing their alignment as fragments of known proteins, PSD-95/SAP90 (clones 1–3) and hdlg/SAP97 (clone 4). Clone 5 cDNA derives from a novel gene (Genbank accession number U32376) with similar sequence and overall structure to PSD-95, and aligns closely with PDZ-1 and PDZ-2 repeats of PSD-95 and hdlg. Numbers in parentheses refer to the number of times each clone was isolated in the screen. The PDZ, SH3 and putative guanylyl kinase (GK) domains are indicated by coloured boxes. White boxes in hdlg represent alternative splice variants⁹. Small numbers refer to the amino-acid residues at the boundaries of domains or cDNA clones. b, Mapping the domains of PSD-95 which mediate binding to Kv1.4C. Various fragments of PSD-95 (shown aligned below the full-length PSD-95 protein) were tested for Kv1.4C binding by semi-quantitative yeast two-hybrid interaction assays, based on the degree of induction by reporter genes HIS3 and β -gal²⁶. HIS3 activity was measured by the percentage of colonies growing on histidine-lacking medium: + + +, >60%; + +, 30–60%; +, 10–30%; –, no significant growth. β -gal activity was determined from the time taken for colonies to turn blue in X-gal filter lift assays²⁶ at room temperature: + + +, <45 min; + +, 45–90 min; +, 90–240 min; –, no significant β -gal activity. c, Amino-acid sequence alignment of PDZ-2 from PSD-95 (ref. 2), hdlg (ref. 9) and clone 5. Identical or similar residues are boxed in black or grey, respectively.

METHODS. Yeast two-hybrid screening^{26,27} was done by using the L40 yeast strain harbouring reporter genes *HIS3* and β -gal under the control of upstream LexA-binding sites. The bait (Kv1.4C) consisted of the C-terminal residues 568–655 of Kv1.4 fused to LexA. About 700,000 clones were screened using a human brain cDNA library (Clontech) constructed in the GAL4-activation-domain vector pGAD10. Eleven positive clones interacting specifically with Kv1.4C (but not with Kv4.2²⁸ or with the NMDA-receptor subunit NR1²⁹) were obtained, yielding multiple isolates of the five cDNAs (a). For domain analysis, different regions of PSD-95 were amplified by PCR with specific primers, and fused to the GAL4-activation domain in pGAD10. These constructs were assayed for interaction with LexA-Kv1.4C by measuring reporter gene activation in the two-hybrid system²⁶. None of the LexA fusions activated reporter gene expression on their own.

[†] To whom correspondence should be addressed.

results). The alignment of these interacting cDNAs reveals an overlap in the N-terminal region of PSD-95 and hdlg/SAP97, which contains three PDZ repeat domains of previously unknown function² (Fig. 1a). Interestingly, only the second PDZ repeat (PDZ-2) of these proteins is found in its entirety in all the isolated interacting clones. The amino-acid sequences of PDZ-2 from PSD-95, hdlg/SAP97, and from clone 5 show extremely high similarity (80–90% identity; Fig. 1c).

Progressive deletion analysis of Kv1.4C from both N- and C-terminal sides revealed that the C-terminal 11 residues (645–655) are sufficient, and the last 4 amino acids (652–655) are essential for PSD-95 interaction. These findings were confirmed both by yeast two-hybrid interaction, as well as by direct *in vitro* binding of bacterial fusion proteins in a filter overlay assay (Fig. 2a, b). The sequence of the C-terminal three amino acids of Kv1.4 (Thr-Asp-Val 655 or TDV) is strikingly conserved in other members of the Shaker⁷ but not Shaw, Shab or Shal subfamilies of voltage-gated K⁺ channels¹¹. We found by yeast two-hybrid and overlay assays that PSD-95, hdlg/SAP97 and clone 5 also bind the C-terminal tails of other Shaker-type subunits Kv1.1, Kv1.2



and Kv1.3 (Fig. 2c), suggesting that the C-terminal TDV sequence may define part of a critical motif for interaction with PDZ-1 or PDZ-2 of PSD-95 family members. This was supported by mutational analysis: substitution of Thr 653 to Ala, or of Val 655 to Ala or Glu abolished interaction with PSD-95 (Fig. 2d). In contrast, various mutations of Ala 650 and Val 651 had no effect, and conservative mutations of Glu 652 to Gln, and of Thr 653 to Ser were also tolerated. Intriguingly, PSD-95 binding activity was retained by the C-terminal sequence ESDV, the exact sequence found at the C terminus of both NR2A and NR2B subunits of the NMDA receptor¹²⁻¹⁴. Taken together, the results indicate that the ~90-amino-acid-long PDZ domains may function as independent protein-binding modules mediating interactions with specific short peptide sequences, analogous to SH2/SH3 domains^{15,16}.

NATURE · VOL 378 · 2 NOVEMBER 1995

FIG. 3 Double immunofluorescence staining of rat cerebellum, showing example of colocalization of Shaker-subfamily K^+ channels and PSD-95 *in vivo*. Kv1.2 immunoreactivity (red) colocalizes with PSD-95 immunoreactivity (green) in the nerve terminal plexuses of basket cells (pinneaus), which form a sheath round Purkinje neuron initial axon segments. M, molecular layer; G, granule cell layer; W, white matter. Scale bar, 100 μ m (left panels), 10 μ m (right panels).

METHODS. Guinea-pig anti-PSD-95 antibodies were raised against a hexahistidine fusion protein of PSD-95 clone 2; these recognized PSD-95 fusion proteins and a ~95K brain protein on immunoblots (not shown). Given the sequence similarity between PSD-95, hdlg/SAP97, and clone 5 in the region of the clone 2 fusion protein, it is possible that these antibodies will crossreact with all three proteins. Rabbit anti-Kv1.2 antibodies, and immunohistochemical methods have been described^{17,19,20}. All primary antibodies were affinity-purified and used at ~1 μ g ml⁻¹ concentration. Immunostaining patterns were visualized by Cy3-conjugated anti-rabbit, and FITC-labelled anti-guinea-pig secondary antibodies (Jackson Labs). Staining patterns for each antibody were specific, as judged by controls with no primary or different primary antibodies, and are in agreement with published distributions of Kv1.2 and PSD-95 (refs 8, 20, 21).

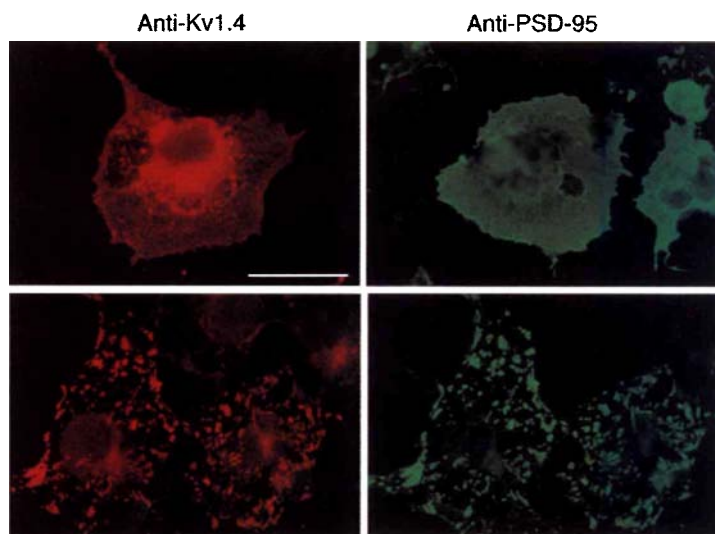
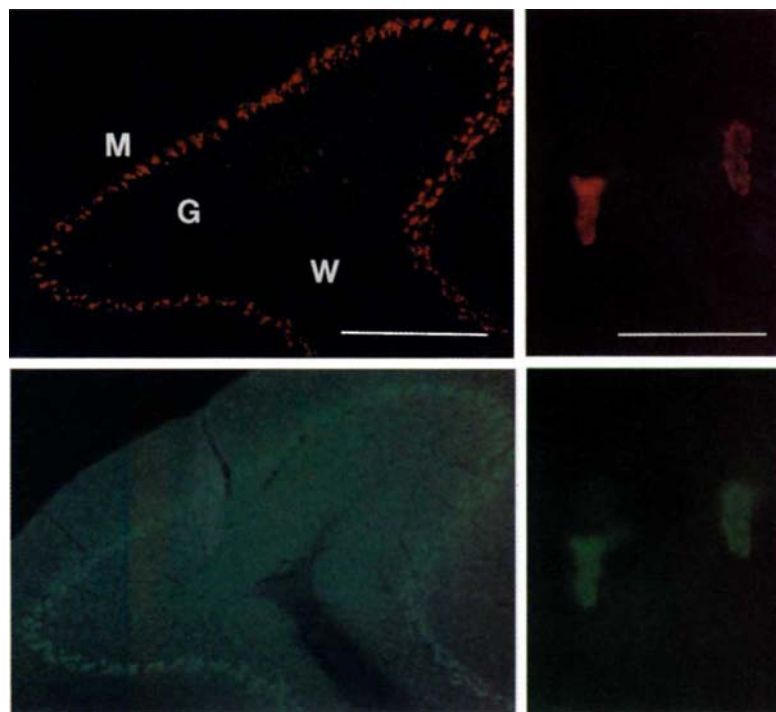


FIG. 4 Clustering of Kv1.4 induced by coexpression of PSD-95 in heterologous cells. COS7 cells were transfected with Kv1.4 alone (upper left), PSD-95 alone (upper right), or with both cDNAs (bottom panels). The distribution of the respective proteins was detected by rabbit anti-Kv1.4 antibodies (red) or by guinea-pig anti-PSD-95 antibodies (green), as indicated. Coexpression results in a remarkable co-clustering of Kv1.4 and PSD-95 immunoreactivities into irregular patches on the surface of transfected cells (seen in essentially all cells that expressed both proteins). Some accumulation of Kv1.4 and PSD-95 can also be seen intracellularly in the perinuclear region (probably in the *trans*-Golgi system) of singly and doubly transfected cells. Similar co-clustering results were also obtained in HEK293 and quail QT-6 fibroblasts. Scale bar, 10 μ m.

METHODS. Kv1.4 and PSD-95 cDNAs were subcloned into mammalian expression vector GW1-CMV (British Biotechnology) and transfected into COS7 cells by the lipofectamine method (Gibco BRL). Cells were fixed with 2% formaldehyde or methanol two days after transfection and double immunofluorescence immunocytochemistry was performed as for Fig. 3. Anti-Kv1.4 antibodies have been described¹⁹.

tion is consistent with direct association of PSD-95 proteins with Shaker-type K^+ channels *in vivo*.

To determine the functional significance of their interaction, PSD-95 and Kv1.4 cDNAs were transiently transfected into COS7 cells, and their protein distributions analysed by double immunofluorescence microscopy. When expressed individually, Kv1.4 is diffusely distributed on the cell surface, whereas PSD-95 is found diffusely throughout the cytoplasm (Fig. 4 and data not shown). When expressed together in the same cell, however, a dramatic redistribution of both proteins occurs. In cotransfected cells, PSD-95 and Kv1.4 immunoreactivities colocalize strikingly on the cell surface in multiple irregularly shaped patches of ~0.5–2.0 μ m diameter (Fig. 4). In contrast, no clustering of the Shal-type voltage-gated K^+ channel Kv4.2 was observed after coexpression with PSD-95 (not shown). Moreover, Kv1.4 and PSD-95 can be coimmunoprecipitated from detergent extracts of cotransfected cells (E.K., unpublished results). The co-clustering of Kv1.4 and PSD-95 is reminiscent of gephyrin–glycine receptor co-clusters in spinal cord neurons²², and of clustering of acetylcholine receptors induced by coexpres-

sion of 43K/rapsyn²³. In contrast to 43K/rapsyn, however, PSD-95 does not form clusters when expressed by itself in heterologous cells.

Taken together with the yeast two-hybrid and *in vitro* binding data, the heterologous coexpression experiments indicate that the clustering of Kv1.4 may be directly mediated by the binding of its C-terminal tail to PSD-95. K^+ channel clustering could potentially be achieved by physical linkage of neighbouring channels (each of which contains four Kv1.4 subunits) via the two binding sites in PSD-95 (PDZ-1 and PDZ-2), or possibly by higher order 'crosslinking' if PSD-95 existed as a multimer. Given the colocalization of these proteins in rat brain, it seems reasonable to suggest that interactions with the PSD-95 family of membrane-associated proteins could underlie the subcellular clustering and anchoring of Shaker-type voltage-gated K^+ channels *in vivo*.

Compared with Shaker-type K^+ channels, the PSD-95 family represent abundant proteins of the brain. It seems likely that PSD-95 and its relatives will have additional functions in neurons, probably by binding to, and organizing the surface distri-

bution of, a variety of integral membrane proteins. Consistent with this idea is the fact that mutations of *discs large*²⁴ (encoding a *Drosophila* homologue of PSD-95) result in profound disorganization of synaptic structure²⁵. It will be of interest to identify additional molecules interacting with PDZ domains of PSD-95, and to investigate how these interactions might influence the functions of the SH3 and guanylate kinase domains of PSD-95 family proteins. □

Received 4 August; accepted 7 September 1995.

1. Froehner, S. C. *Rev. Neurosci.* **16**, 347–368 (1993).
2. Cho, K.-O., Hunt, C. A. & Kennedy, M. B. *Neuron* **9**, 929–942 (1992).
3. Willott, E. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7834–7838 (1993).
4. Adams, M. E. et al. *Neuron* **11**, 531–540 (1993).
5. Li, M., Jan, Y. N. & Jan, L. Y. *Science* **257**, 1225–1230 (1992).
6. Hoshi, T., Zagotta, W. N. & Aldrich, R. W. *Science* **250**, 533–538 (1990).
7. Stühmer, W. et al. *EMBO J.* **8**, 3235–3244 (1989).
8. Kistner, U. et al. *J. biol. Chem.* **268**, 4580–4583 (1993).
9. Lue, R. A., Marfatia, S. M., Branton, D. & Chishti, A. H. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9818–9822 (1994).
10. Müller, B. M. et al. *J. Neurosci.* **15**, 2354–2366 (1995).

11. Chandy, K. G. & Gutman, G. A. *Handbook of Ligand and Voltage-gated Ion Channels* 1–71 (CRC, Boca Raton, FL, 1995).
12. Ishii, T. et al. *J. biol. Chem.* **268**, 2836–2843 (1993).
13. Kutsuwada, T. et al. *Nature* **358**, 36–41 (1992).
14. Monyer, H. et al. *Science* **256**, 1217–1221 (1992).
15. Pawson, T. *Nature* **373**, 573–580 (1995).
16. Cohen, G. B., Ren, R. & Baltimore, D. *Cell* **80**, 237–248 (1995).
17. Sheng, M., Liao, Y. J., Jan, Y. N. & Jan, L. Y. *Nature* **365**, 72–75 (1993).
18. Wang, H., Kunkel, D. D., Martin, T. M., Schwartzkroin, P. A. & Temple, B. L. *Nature* **365**, 75–79 (1993).
19. Sheng, M., Tsaur, M.-L., Jan, Y. N. & Jan, L. Y. *Neuron* **9**, 271–284 (1992).
20. Sheng, M., Tsaur, M.-L., Jan, Y. N. & Jan, L. Y. *J. Neurosci.* **14**, 2408–2417 (1994).
21. Wang, H., Kunkel, D. D., Schwartzkroin, P. A. & Tempel, B. L. *J. Neurosci.* **14**, 4588–4599 (1994).
22. Kirsch, J., Wolters, L., Triller, A. & Betz, H. *Nature* **366**, 745–748 (1993).
23. Phillips, W. D. et al. *Science* **251**, 568–570 (1991).
24. Woods, D. F. & Bryant, P. J. *Cell* **66**, 451–464 (1991).
25. Lahey, T., Gorczyca, M., Jia, X.-X. & Budnik, V. *Neuron* **13**, 823–835 (1994).
26. Bartel, P. L., Chien, C.-T., Sternberg, R. & Fields, S. *Using the 2-Hybrid System to Detect Protein-Protein Interactions in Cellular Interactions in Development: A Practical Approach* 153–179 (Oxford University Press, 1993).
27. Lee, J. E. et al. *Science* **268**, 836–844 (1995).
28. Baldwin, T. J., Tsaur, M.-L., Lopez, G. A., Jan, Y. N. & Jan, L. Y. *Neuron* **7**, 471–483 (1991).
29. Moriyoshi, K. et al. *Nature* **354**, 31–37 (1991).

ACKNOWLEDGEMENTS. We thank C.-T. Chien for teaching us the two-hybrid system, J. U. Kim for technical support, and E. Aidonis for help with the manuscript. Y.N.J. is an investigator and M.S. is an assistant investigator of the Howard Hughes Medical Institute.

Cloning of a new cytokine that induces IFN- γ production by T cells

Haruki Okamura*, Hiroko Tsutsui†, Toshinori Komatsu‡, Masuo Yutsudo§, Akira Hakura§, Tadao Tanimoto||, Kakuji Torigoe||, Takanori Okura||, Yoshiyuki Nukada||, Kazuko Hattori||, Kenji Akita||, Motoshi Namba||, Fujimi Tanabe||, Kaori Konishi||, Shigeharu Fukuda|| & Masashi Kurimoto||

* Department of Bacteriology and ‡ Department of Immunology and Medical Zoology, Hyogo College of Medicine, 1-1, Mukogawa-cho, Nishinomiya, Japan

† Toneyama Institute for Tuberculosis, Osaka City University, Toneyama-cho, Toyonaka, Japan

§ Institute for Microbial Disease, Yamada-kami, Suita, Japan

|| Fujisaki Institute, Hayashibara Biochemical Laboratories, Hayashibara Co. Inc., Okayama, Japan

THE mechanism underlying the differentiation of CD4⁺ T cells into functionally distinct subsets (Th1 and Th2) is incompletely understood^{1–3}, and hitherto unidentified cytokines may be required for the functional maturation of these cells⁴. Here we report the cloning of a recently identified IFN- γ -inducing factor (IGIF) that augments natural killer (NK) activity in spleen cells^{5,6}. The gene encodes a precursor protein of 192 amino acids and a mature protein of 157 amino acids, which have no obvious similarities to any peptide in the databases. Messenger RNAs for IGIF and interleukin-12 (IL-12) are readily detected in Kupffer cells and activated macrophages. Recombinant IGIF induces IFN- γ more potently than does IL-12, apparently through a separate pathway. Administration of anti-IGIF antibodies prevents liver damage in mice inoculated with *Propionibacterium acnes* and challenged with lipopolysaccharide, which induces toxic shock. IGIF may be involved in the development of Th1 cells and also in mechanisms of tissue injury in inflammatory reactions.

Interferon- γ -inducing factor (IGIF) was purified from the livers of mice treated with the bacterium *P. acnes* and subsequently challenged with lipopolysaccharide (LPS)⁶ to induce toxic shock. Several peptide fragments were obtained from the purified factor by trypsin digestion, and degenerate oligonucleotide primers were synthesized based on the sequence of the two fragments.

Using these primers, a partial complementary DNA fragment was cloned from messenger RNA extracted from the liver of these mice by reverse transcription-polymerase chain reaction (RT-PCR). The amplified cDNA fragments were used as probes for screening a cDNA library prepared from the liver of the *P. acnes*-injected, LPS-challenged mice. A full-length clone obtained from the cDNA library encoded a precursor polypeptide of 192 amino acids (Fig. 1a): no homologous peptides were found in the EMBL, GenBank or PIR databases (our nucleotide-sequence data will appear in the GSDB, DDBJ, EMBL and NCBI nucleotide sequence databases under the accession number D49949). The precursor polypeptide contained an unusual leader sequence composed of 35 amino acids and no *N*-glycosylation sites or putative hydrophobic signal peptide. An extract from *Escherichia coli* transformed with the cDNA co-stimulated IFN- γ production by plastic-nonadherent spleen cells in the presence of anti-CD3 monoclonal antibody or a suboptimal dose of concanavalin A. The recombinant peptide (rIGIF) was extensively purified by column chromatography: SDS-PAGE showed that its activity coincided with a species of relative molecular mass 18–19K (Fig. 1b).

Recombinant IGIF markedly stimulated IFN- γ production in established Th1 cells (Fig. 2a) in the presence of anti-CD3 monoclonal antibody. The levels of IFN- γ induced by rIGIF were much higher than those induced by IL-12 (Fig. 2b). There was an interesting synergistic effect on IFN- γ production between IGIF and IL-12 (Fig. 2c). Similar results were obtained when spleen T cells were enriched using nylon wool, but established Th1 clones were used to avoid the effects of endogenous cytokines produced by heterogeneous cell populations. Addition of neutralizing murine anti-IL-12 antibody (from G. Trinchieri) to the culture did not inhibit the production of IFN- γ induced by IGIF (Fig. 2a), indicating that IGIF does not depend on IL-12 for its activity. Moreover, addition of anti-IGIF antibody did not inhibit IFN- γ induction by IL-12 (Fig. 2b), confirming that these cytokines can act independently of one another. As well as inducing IFN- γ production, IGIF also stimulated the proliferation of T cells (Fig. 2d) and augmented the lytic activity of spleen cells on YAC-1 cells (Fig. 2e). This cytokine thus apparently shares some of its biological activities with IL-12, although it is a different molecule².

We next determined which cells produce IGIF. We found that IGIF mRNA was highly expressed in untreated or *P. acnes*-elicited Kupffer cells and that LPS had almost no effect on this expression (Fig. 3). A moderate amount of IL-12 p40 mRNA (encoding IL-12 of *M_r* 40K) was also detected in *P. acnes*-elicited Kupffer cells, and this was augmented by LPS after a lag period

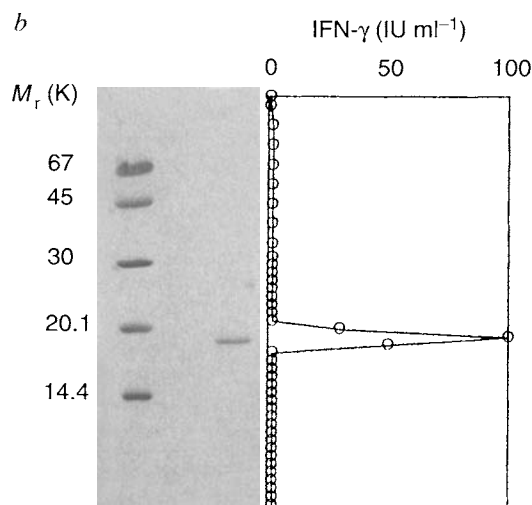
FIG. 1 a, Nucleotide sequence of IGIF cDNA and predicted amino-acid sequence. An IFN- γ -inducing factor was purified from the livers of mice injected with *P. acnes* (3 mg; i.p. for 1 w), then challenged with LPS (10 μ g, i.v.) 2 h before being killed. The factor was purified by column chromatography on hydrophobic phenyl-Sepharose CL-4B, anion exchange on DEAE-Sepharose CL-6B, chromatofocusing using Mono P, ultrafiltration through Superdex-75, and reverse-phase chromatography by Protein C4. The ability of IGIF to induce IFN- γ production in the presence of a suboptimal dose of concanavalin A (0.5 μ g ml $^{-1}$) was assayed using spleen plastic-nonadherent cells of C3H/HeJ mice, and IFN- γ was titrated by ELISA. The partially purified protein was resolved by SDS-PAGE and further purified by extracting the IFN- γ -inducing species (20 μ g) from the sliced gel and digesting it with trypsin (2 μ g ml $^{-1}$ at 37 °C for 18 h). The N-terminal amino-acid sequence of each fragment was determined using Perkin-Elmer protein sequencer type 473 A. Mixtures of DNA oligomers (20 bases) believed to correspond to the amino-acid sequences of the two fragments of interest were constructed by an automated DNA synthesizer (Applied Biosystems) and used as primers (5'-ATRTCTCDATRTTTCNGG-3', and 5'-TTYGARGAYATGACNGAYAT-3', respectively). Total RNA was extracted from 3 g of livers of mice with LPS-induced shock and amplified by RT-PCR. Amplified DNA fragments were probed by Southern blotting and ligated into plasmid pT7 blue T-Vector (Novagen); the recombinant DNA (MPCR-6) was used to transform competent *E. coli* NoVa Blue (Novagen). DNA that hybridized with the labelled probe was isolated and used as a hybridization probe to screen a cDNA clone containing the complete IGIF coding region. A cDNA library was constructed using a cloning kit (Amersham, System Plus) from poly(A $^{+}$)RNA prepared from livers of mice with LPS-induced shock, and integrated into the *Eco*RI site of a λ gt10 phage vector. Positive plaques were screened by hybridization using 32 P-labelled MPCR-6 as probe, and a 0.9-kb insert clone was sequenced (DNA sequencer model 373A; ABI). The predicted amino-acid sequences underlined were identified by sequencing of fragments of purified IGIF. The N-terminal amino acid is indicated by an asterisk. b, Analysis of rIGIF by SDS-PAGE. An IGIF cDNA fragment was cloned into the pKK223-3 expression vector (Pharmacia) and transfected into *E. coli*. Recombinant IGIF was purified from the extract of the transformant grown in L-broth by column chromatography using phenyl-Sepharose CL-4B, DEAE Sepharose and Superdex 75. This purified protein was

a

```

1 ATCACAGGCACAGCTGGACCTGGTGGGGTCTCTCTGTTCCATGCTTTCTGGACTCCTGCCTGCTGGCTGGAGCTGCT
81 GACAGGCCTGACATCTTCTGCAACCTCCAGCATCAGGACAAAGAAAGCCGCTCAAACTTCCAAATCACTTCTCTGGCCAGGAACA
171 ATGGCTGCCATGTCAGAAGACTCTTCGCTCAACTTCAAGGAAATGATGTTTATGACAACACGCTTTACTTTTATACCTGAAGAAATGGA
M A A M S E D S C V N F K E M M F I D N T L Y F I P E E N G
1 10 20 30
261 GACCTGGAATCAGACAACCTTTGGCCGACTTCACTGTACAACCGCAGTAATACGGAATATAAATGACCAAGTTCTCTTCGTTGACAAAAGA
D L E S D N F G R L H C T T A V I R N I N D Q V L F V D K R
* 40 50 60
351 CAGCCTGTGTTTCGAGGATATGACTGATATGATCAAAAGTCCAGTGAACCCAGACAGACTGATAATATACATGTACAAAGACAGTGAA
Q P V F E D M T D I D Q S A S E P Q T R L I I Y M Y K D S E
70 80 90
441 GTAAGAGGACTGGCTGTGACCTCTCTGTGAAGGATAGTAAATGTCTACCTCTCTGTAAGAACAAGATCATTTCTTTGAGGAAATG
V R G L A V T L S V K D S K M S T L S C K N K I I S F E E M
100 110 120
531 GATCCACCTGAAATATTGATGATATACAAAGTGTCTCATATTCTTTCAGAAACGTGTTTCCAGGACACACAAGATGGAGTTTGAATCT
D P P E N I D D I Q S D L I F F Q K R V P G H N K M E F E S
130 140 150
621 TCACTGTATGAAGGACACTTCTCTGCTTGCCAAAAGGAGATGATGCTTTCAAACTCATTTCTGAAAAAAGAGTGAATGGGGATAA
S L Y E G H F L A C Q K E D D A F K L I L K K K D E N G D K
160 170 180
711 TCTGTAATGTTCACTCTCACTAACTACATCAAGTTAGTGGGGAGGGTTTGTGTTCCAGAAAGATGATTAGCACACATGCGCCTTGTG
S V M F T L T N L H Q S
190
801 ATGACCTCGCTGTATTTCCTAATACAGAATACCCGAGGCTGCATGATTATAGAGTAAACACGTTTATTGTAAAAA

```



applied to SDS-PAGE and one lane was stained with CBB (left). Proteins were extracted from the gel slices of a similar lane and assayed for IGIF activity. The IGIF activity extracted is depicted alongside the corresponding position on the gel (right).

of several hours. Neither IGIF mRNA nor IL-12 p40 mRNA could be detected in peritoneal adherent cells, but they were strongly expressed in *P. acnes*-induced peritoneal exudate cells. Thus, IGIF and IL-12 p40 both seem to be expressed in activated macrophages. However, IGIF activity could only be detected in the culture supernatant when these cells were stimulated with LPS (Fig. 3).

To investigate the physiological functions of IGIF *in vivo*, we tested the effect of neutralizing anti-IGIF antibody on LPS-induced liver injury. As shown previously, treatment of mice with *P. acnes* and subsequent challenge with LPS causes serious hepatic injury similar to fulminant hepatitis^{7,8}. Although IL-12 and other mediators have been implicated in shock-induced death of mice injected with bacillus Calmette-Guérin (BCG) and LPS⁹, the mechanism of this pathological effect is unclear. Because athymic nude mice are susceptible to hepatic injury but resistant to the lethal shock caused by these treatments, we used nude mice to simplify interpretation of the results. As shown in Fig. 4a, a focal necrosis developed in mice treated with control

rabbit preserum. Necrosis was accompanied by bleeding, which resulted from damage to sinusoidal endothelial cells. We found that administration of neutralizing anti-IGIF antiserum instead of the preserum completely prevented liver injury in these mice (Fig. 4b). Our observations were confirmed by determining serum levels of the transaminases GPT and GOT (Fig. 4c). We conclude from both our histological and biochemical results that IGIF is a critical factor in the induction of liver injury caused by endotoxin exposure.

Cytokines produced during inflammatory reactions caused by initial contact of antigens with host cells are thought to be responsible for the resulting immune responses through regulation of the development of Th subsets of CD4 $^{+}$ T cells¹. IGIF may turn out to be one of these regulatory cytokines. Elucidation of the regulatory mechanisms that control the production of this cytokine and of IL-12 in macrophages will also be important for the understanding of the development of cellular immunity. This new cytokine augments the cytotoxic effect of spleen cells and is profoundly involved in tissue injury in serious hepatitis

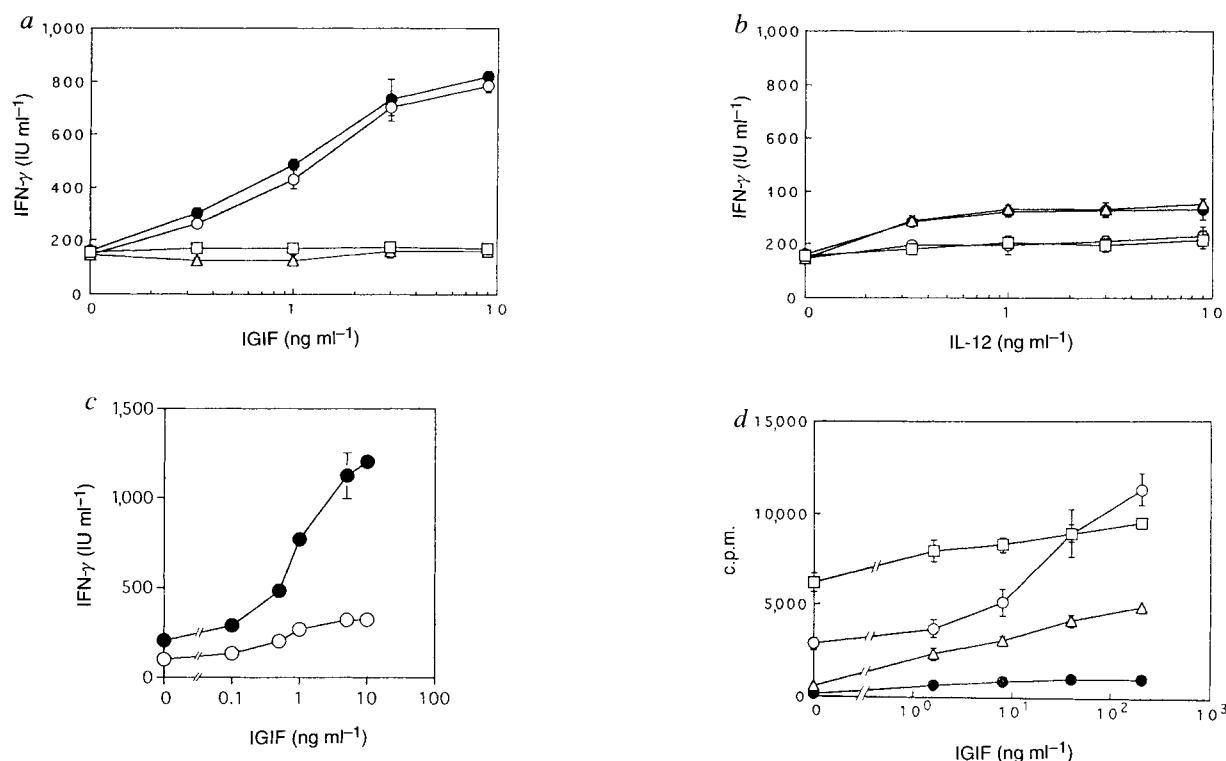
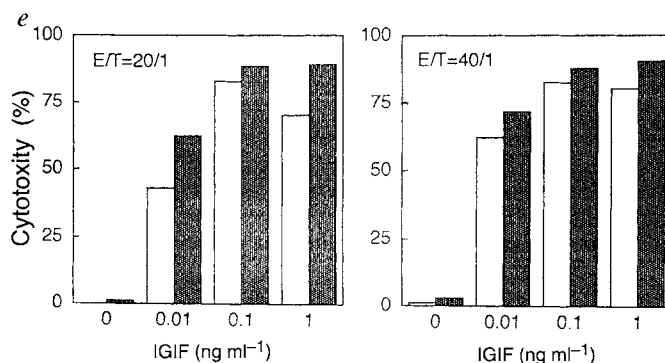


FIG. 2 Biological activities of rIGIF. *a*, Effects of rIGIF on the production of IFN- γ by Th1 cells. Purified rIGIF was added to the cultures of Th1 cells (2×10^4 cells in 150 μ l of RPMI 1640 medium supplemented with 10% FCS per well) in microtiter plates coated with anti-CD3 mAb (5μ g ml⁻¹) with (○), or without (●) anti-muIL-12 mAb (50μ g ml⁻¹), and the culture supernatants were assayed for IFN- γ activity after 48 h using the ELISA method. Culture supernatants added with anti-IGIF Ab (25μ g ml⁻¹) instead of anti-IL-12 mAb (△) or with the combination of anti-IGIF and anti-IL-12 (□) were assayed for IFN- γ produced. *b*, Effect of IL-12 on IFN- γ induction by Th1 cells. Murine rIL-12 was produced by the cultures of CHO cells transfected with cDNA for IL-12 p35 and p40 proteins according to the described methods¹⁰. Murine rIL-12 was purified by anion exchange and hydroxyapatite column chromatographies. IL-12 was added to the cultures of Th1 cells at the indicated doses in microtiter plastic plates coated with anti-CD3 mAb with (△) or without (●) anti-IGIF Ab (25μ g ml⁻¹), and the IFN- γ was assayed as described in Fig. 2a. Culture supernatants added with anti-IL-12 mAb (○) or with combination with anti-IGIF Ab and anti-IL-12 mAb (□) instead of anti-IGIF Ab were also assayed for IFN- γ . *c*, Synergistic effect of IGIF and IL-12 on IFN- γ production. Th1 cells (2×10^4 cells per well) were incubated with various concentrations of IGIF in the absence (○) or in the presence (●) of constant dose of IL-12 (20μ g ml⁻¹) in the microtiter plates coated with anti-CD3 mAb as described in *a*. Culture supernatants were assayed for IFN- γ after 24 h. *d*, The effect of rIGIF on T-cell proliferation. Recombinant IGIF was serially diluted and added to cultures of C3H/HeJ mouse plastic nonadherent spleen cells (1.5×10^5 cells per well) in the absence (●), or in the presence of anti-CD3 mAb (1μ g ml⁻¹, □), Con A (0.5μ g ml⁻¹, ○), or human rIL-2 (10μ g ml⁻¹, △). The cultures



were pulsed with 0.5 μ Ci of ³H-thymidine during the final 24 h of a 72 h incubation. The amount of radioactivity incorporated was measured using a Packard Direct Beta Counter (Matrix 96, CT, U.S.A.). *e*, The effect of rIGIF on cytotoxic activity. Spleen cells from BDF1 mice were incubated for 3 d with rIGIF at various dilutions in the absence (open bar) or in the presence (closed bar) of IL-2 at the indicated concentrations. The cells were washed and added to YAC-1 cells (10^5 cells ml⁻¹) that had been labelled with Na²⁵¹Cr₂O₄ at various effector/target cell ratios. Cytotoxicity is expressed as the percent lysis determined by conventional methods.

caused by endotoxin. The nature of the involvement needs to be clarified so that therapeutic strategies can be developed against various inflammatory diseases. It may influence the development

of tumours and allergies and the pathogenesis of diseases such as pulmonary tuberculosis, leprosy, fulminant hepatitis and AIDS. □

Received 11 April; accepted 4 October 1995.

1. Mosmann, T. R. & Coffman, R. L. A. Rev. Immun. **7**, 145–173 (1989).
2. Scott, P. Science **260**, 496–497 (1993).
3. Le-Gros, G. S., Ben-Sasson, S. Z., Seder, R., Finkelman, F. D. & Paul, W. E. J. exp. Med. **172**, 921–929 (1990).
4. Locksley, R. M. Proc. natn. Acad. Sci. U.S.A. **90**, 5879–5880 (1993).
5. Nakamura, K., Okamura, H., Nagata, K., Komatsu, T. & Tamura, T. Infect. Immun. **61**, 64–70 (1993).
6. Okamura, H. et al. Infect. Immun. (in the press).
7. Ferluga, J. & Allison, A. C. Lancet **661**–662 (1978).
8. Tsutsui, H., Mizoguchi, Y. & Morisawa, S. Hepato-gastroenterology **39**, 553–559 (1992).

9. Wysocka, M. et al. Eur. J. Immun. **25**, 672–676 (1995).
10. Karasuyama, H., Tohyama, N. & Tada, T. J. exp. Med. **166**, 13–25 (1989).
11. Eyhorn, S. et al. J. Hepatol. **6**, 23–35 (1988).
12. Chomczynski, P. & Sacchi, L. Anal. Biochem. **162**, 156–159 (1987).
13. Wynn, T. A., Eltoun, I., Oswald, I. P., Cheever, A. W. & Sher, A. J. Exp. Med. **179**, 1551–1561 (1994).

ACKNOWLEDGEMENTS. We thank N. Kawada, K. Kohno, I. Okamoto, K. Nakanishi, K. Matsui and S. Kashiwamura for preparing Kupffer cells, Th1 cells and murine rIL-12, and for discussion and support.

FIG. 3 The expression of IGIF in *P. acnes*-elicited Kupffer cells (A) and *P. acnes*-induced peritoneal cells (B). A, Kupffer cells were isolated from BALB/c mice that had been treated with *P. acnes* 7 d previously, as described¹¹. Non-parenchymal cells were isolated using pronase-collagenase and the elutriated fractions were collected at a pump speed of 36.2–47.5 ml min⁻¹ in a total of 150 ml of medium using CR20B2 (Hitachi, Tokyo). The cells were washed, resuspended in RPMI supplemented with 20% FCS, plated on plastic plates (9 cm), and incubated for 24 h. The plastic adherent cells (3×10^6 cells per ml) were additionally incubated with LPS ($1 \mu\text{g ml}^{-1}$) for the indicated hours. The cells were homogenized with 0.5 ml of denaturing solution (4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0, 0.1 M mercaptoethanol, 0.5% N-lauroylsarcosine) and the RNA was extracted according to established methods¹². The expression of IGIF (a), IL-12 (p40, b) and $\beta 2$ -microglobulin (c) was estimated by RT-PCR. The RNA was reverse-transcribed and amplified using RT-PCR kits (Perkin-Elmer, Norwalk, CT). The thermocycle conditions were 35 cycles of 95 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min, for denaturing, annealing and extension, respectively. The primer sequences were, 5'-ACTGTACAACCGCAGTAATACGG-3' and 5'-AGTGAACATTACAGATTATCCC-3' for IGIF, 5'-CGTGCTATGGCTGGTGCAAG-3' and 5'-GAACACATGCCCATTTGCTG-3' for p40 of IL-12, and 5'-TCAGATCTGCTTCAGCAA-3', and 5'-CATGTCTCGGTCCCAGGTGA-3' for $\beta 2$ microglobulin¹³, respectively. Amplified DNAs were resolved by means of 2% agarose gel electrophoresis and stained with ethidium bromide (a, IGIF; b, p40; c, $\beta 2$ microglobulin). 1, Kupffer cells from untreated mice. 2, Kupffer cells from untreated mice and incubated with LPS for 6 h. 3–8, Kupffer cells from *P. acnes*-injected mice, incubated with LPS for 0 h, 1 h, 2 h, 3 h, 6 h and 24 h. The IGIF activity in the culture supernatant incubated with (○) or without (●) LPS was also assayed by measuring its ability to induce IFN- γ (right). B, *P. acnes*-induced peritoneal exudate cells were prepared from mice injected (i.p.) with heat-killed *P. acnes* (3 mg) 4d previously, and the adherent cells were incubated with LPS. RT-PCR and IGIF assay were carried out as described for Kupffer cells. 1, resident peritoneal adherent cells. 2, resident peritoneal adherent cells, incubated with LPS for 6 h. 3–8, *P. acnes*-induced peritoneal adherent cells, incubated with LPS for 0 h, 1 h, 4 h, 8 h, 12 h and 24 h.

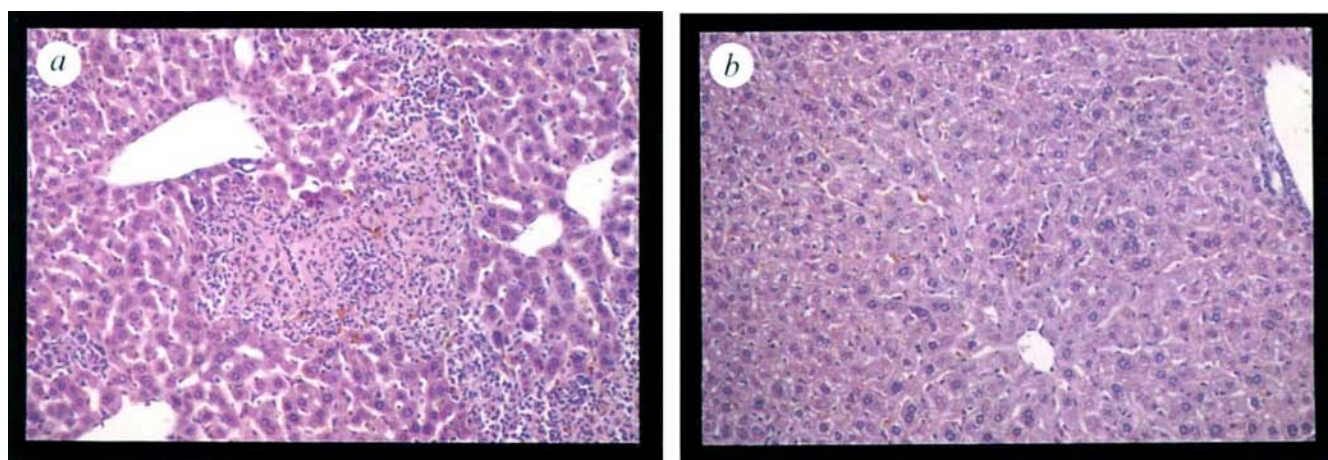
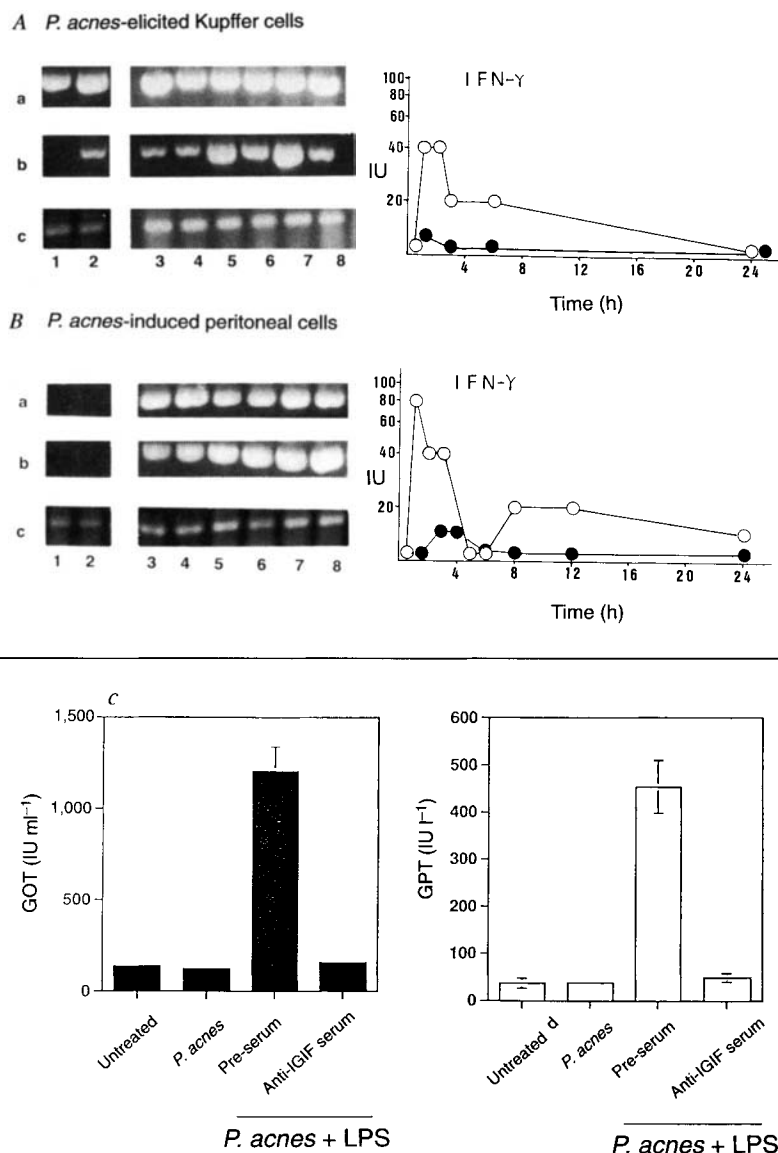


FIG. 4 Neutralizing anti-IGIF antibody completely inhibited endotoxin-induced hepatic injury. Male BALB/c *nu/nu* mice (5 w old) were injected (i.p.) with 40 μl of control rabbit pre-serum (a) or with the same volume of neutralizing anti-IGIF antiserum raised in a rabbit (b). After 30 min, they were injected with 1 mg of *P. acnes* (i.v.). At days 2, 4 and 6, they were additionally injected with the same sera, respectively. These mice

were challenged with LPS on day 7, and the sera and the liver specimens were sampled after 24 h. The liver specimens were fixed in 3.5% formaldehyde in PBS and stained with haematoxylin and eosin (a, b). Alanine transaminase (GPT) and aspartate aminotransferase (GOT) in the sera were also measured (c).

A eukaryotic transcriptional repressor with carboxypeptidase activity

Gong-Ping He, Aleixo Mulse, Audrey Wu LI & Hyo-Sung Ro*

Department of Biochemistry, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, B3H 4H7, Canada

ADIPOCYTE differentiation involves the transcriptional activation of several genes in triglyceride metabolism, including the adipose *P2* (*aP2* or 422) gene that encodes the adipocyte lipid-binding protein ALBP¹⁻³. Within the mouse *aP2* promoter region, the AE-1 sequence functions as either a positive or a negative element in the regulation of *aP2* gene expression⁴⁻⁸. The AE-1 sequence is the binding site for the positive murine (3T3) adipocyte factor C/EBP- α ^{5,6}, several human preadipocyte factors⁷, and a 3T3 preadipocyte factor(s) that has been implicated as a repressor of *aP2* gene expression^{4,5,7,8}. Here we report the cloning of new complementary DNAs that encode the 3T3 preadipocyte factor (termed AEBP1), and demonstrate that AEBP1 expression is abolished during adipocyte differentiation. Furthermore, we show that an activity of a carboxypeptidase associated with AEBP1 is important in the transcriptional repression function of AEBP1. Thus AEBP1 might represent a new type of transcription factor that regulates transcription by cleavage of factors involved in transcription.

AEBP1 cDNAs were cloned, and the sequence of one clone revealed a 719-residue open reading frame (ORF) encoding a polypeptide of estimated relative molecular mass (M_r) 79,000 (79K). The carboxy-terminal sequence of AEBP1 consists of several domains, including basic, acidic and Ser, Thr, Pro-rich sequences (Fig. 1a). In addition, an amino-terminal region of AEBP1 shares about 35% amino-acid identity with a domain in a *Dictyostelium discoideum* lectin, discoidin I. Discoidin I-like domain (DLD), which has been proposed to have cell adhesion properties, is also found in other proteins (Fig. 1b).

Northern blot experiments indicate that AEBP1 messenger RNA (~4.0 kb) decreases in abundance during adipocyte differentiation (Fig. 2a). Western blot analysis detected two cellular proteins of 45K and 30K in 3T3-L1 preadipocyte cells (Fig. 2b, lane 1), but not in adipocytes (lane 2). These proteins could be proteolytic fragments of AEBP1, as the preadipocyte extracts prepared in more stringent conditions contained only the expected full-length AEBP1 protein (lanes 5 and 6).

Because expression of AEBP1 is abolished during adipocyte differentiation, it could be a negative regulator involved in the repression of *aP2* gene expression in preadipocytes. As shown in Fig. 3a, the recombinant AEBP1 protein (Fig. 4a) was able to bind specifically to the AE-1 sequence. Furthermore, AEBP1 represses the expression of chloramphenicol acetyltransferase (CAT) gene which is driven by the *aP2* promoter, containing the AE-1 site (Fig. 3b). The results demonstrate that the repression activity of AEBP1 is specific and is not due to transcriptional 'squenching' or to a general nonspecific shutdown of the RNA polymerase II machinery.

The most striking feature of AEBP1 is a region (amino acids 136-551) that is very similar (40% identity) to almost the entire sequence of the regulatory type carboxypeptidases (CPs) (Fig. 1c). To determine whether AEBP1 is a functional carboxypeptidase, we purified recombinant AEBP1 protein (His-AEBP1; Fig. 4a, line 1), and showed that it did indeed exhibit CP activity (Fig. 4b). However, two frameshift mutations with a truncation in the CP homology region (His-AEBP1 Δ Sac and His-AEBP1 Δ Bam) abolished CP activity (Fig. 4a, lines 3 and 4). Four in-frame deletion mutations near the N terminus of the

CP homology region were also analysed (Fig. 4a, lines 16-19). Two small deletions (Δ Bal-11 and Δ Bal-12; lines 16 and 17) did not abolish the carboxypeptidase activity (Fig. 4b), but the other two deletion mutations (Δ Bg/Ba and Δ Bal-13; lines 18 and 19) did abolish it (Fig. 4b). These results show AEBP1 as a new carboxypeptidase, and establish the importance of the CP homology region for its enzymatic activity.

Because AEBP1 carries out transcriptional repression and has CP activity, we investigated whether the CP activity is involved in the repression function, by analysing a set of deletion and frameshift mutations in the CP homology region. The C-terminal frameshift mutation (His-AEBP1 Δ Sty) did not abolish CP activity (Fig. 4a, line 2), indicating that, in addition to C-terminal non-CP homology region, the C-terminal 37 amino acids (residues 515-551) in the CP homology region are dispensable for CP activity. Similarly, the results for in-frame deletion and frameshift mutations summarized in Fig. 4a (lines 5-12) indicate that the C-terminal 259 amino acids (residues 461-719) of AEBP1 are dispensable for CP activity. Further deletion of the CP homology region (Δ Hic; line 13) resulted in abolition of CP activity (Fig. 4b). These results indicate that residues between positions 429 and 460 in the CP homology region are necessary for CP activity. One small deletion mutation in this region (Δ Bal-9; Fig. 4a, line 14) did not abolish the CP activity (Fig. 4b), but another mutation (Δ Bal-10; line 15) resulted in drastic reduction of CP activity (Fig. 4b). These results suggest that the C-terminal 105 amino acids (residues 447-551) of the CP homology region are not required for CP activity.

To examine the effects of these mutations on transcriptional repression activity, AEBP1 chimaeric proteins with an added DNA-binding specificity were tested in a heterologous system for transcription effects. The cDNA encoding full-length AEBP1 (except for the N terminal two amino acids) was cloned into the GAL4 fusion vector pG4 (ref. 9). The fusion protein (G4-AEBP1) contains the DNA-binding domain (amino acids 1-147) of GAL4 at its N terminus, tethering the fusioin protein to a promoter containing GAL4-binding sites (Fig. 4a). When pG4-AEBP1 was cotransfected with the target gene (pGALTKCAT¹⁰), CAT activity was significantly decreased (Fig. 4c, lanes 1-3), but when tested with pTKCAT (without the GAL4-binding sites) the G4-AEBP1 fusion protein had no effect on CAT expression (lanes 6-8). Therefore, repression by G4-AEBP1 requires localization to the promoter region. The GAL4 frameshift mutant forms of AEBP1 (Δ Sac or Δ Bam), which did not exhibit CP activity, completely lacked the

FIG. 1 Sequence analysis of AEBP1. a, The cDNA and deduced amino acid sequences of AEBP1 clones. The underlined stop codon at the 5' untranslated region (UTR) is in the same reading frame as the first methionine codon at nucleotide 160. A putative polyadenylation signal (AATAAA) in the 3' UTR is underlined. The potentially interesting primary structure motifs highlighted include a region (amino acids 622-656) that is rich in Ser, Thr and Pro (STP-like sequences¹⁷) and which is flanked by a basic (amino acids 598-619, 46% arginine) and an acidic (amino acids 660-699, 58% glutamic acid) region. b, Amino acid sequence alignments of the N-terminal domain of AEBP1 with discoidin I (Disc)¹⁸, the A5 protein¹⁹, the receptor tyrosine kinase (DDR)²⁰, a milk-fat globule protein (MFG-E8)²¹, and the coagulation factors V (CF-V)²² and VIII (CF-VIII)²³ with ~35% sequence identity. c, Amino acid sequence alignments of AEBP1 with regulatory carboxypeptidases (CPH/E²⁴, CPN²⁵ and CPM²⁶). Amino acids thought to be important in the active centres of bovine CPA and B are marked as follows: hash, Zn-binding; circle, catalytic; asterisk, substrate-binding. The vertical lines and dots indicate identical and similar amino acids, respectively; broken lines indicate gaps introduced to optimize the alignments.

METHODS. To clone a cDNA encoding a protein that interacts at the AE-1 site (nucleotides -159 to -125 of the *aP2* gene^{5,7}) in 3T3 preadipocytes, we expressed cDNAs from a 3T3-L1 preadipocyte cell library with a Uni-Zap XR vector (Stratagene) and screened for an expression clone by the Affinity Screening procedure²⁷ with random concatamers of the AE-1 sequence. Three independent phage plaques produced fusion proteins interacting specifically with the AE-1 sequence. Further analysis indicated that one cDNA clone encoded mRNA whose expression was downregulated during adipocyte differentiation (see below). This partial cDNA (~0.7 kb) clone was used as a probe to isolate a full-length cDNA from the same library. One clone contained a ~2.5 kb cDNA, and this cDNA was completely sequenced. The nucleotide sequence data for AEBP1 can be found in the EMBL nucleotide sequence databases under the accession number X80478.

* To whom correspondence should be addressed.

a

1 *EcoR* I Adaptor

gaatcccttactctcgtctgaaccactagcccacagcgggtgtgtcatctgcagagtgcccacctattgggaggagt
81
cacacgcgattgaggacaaccagatccgtgctctccatgctgcgccacggcctcggagccagcgggcccagctcaac

Nae I

160/1
ATG CAG GCT GGT GCC AAT GAA GAT GAC TAC TAT GAC GGG GCA TGG TGT GCT GAG GAC GAG
Met gln ala gly ala asn glu asp tyr thr asp gly ala trp cys ala gly asp glu
220/21
TCG CAG ACC CAG TGG ATC GAG GTG GAC ACC CGA AGG ACA ACT CGG TTC ACG GGC GTC ATC
ser gln thr gln trp ile glu val asp thr arg arg thr thr arg phe thr gly val ile
280/41
ACT CAG GGC CGT GAC TCC AGC ATC CAT GAC GAC TTC GTG ACT ACC TTC TTT GTG GGC TTC
thr gln gly arg asp ser ser ile his asp phe val thr thr phe phe val gly phe
340/61
AGC AAT GAC AGC CAG ACC TGG GTG ATG TAC AAT GAT GGC TAC GAG GAA ATG ACC TTC TAT
ser asn asp ser gln thr trp val met tyr thr asn gly tyr glu glu met thr phe tyr
400/81
GGA AAT GTG GAC AAG GAC ACA CCT GTG CTG AGC GAG CTC CCG GAG GCA GTT GTG GCC CGT
gly asn val asp lys asp thr pro val leu ser glu leu pro glu pro val val ala arg
460/101
TTC ATC CGC ATC TAT CCA CTC ACC TGG AAC GGT AGC CTG TGC ATG CGC CTG GAG GTG CTA
phe ile arg ile tyr pro leu thr trp asn gly ser leu cys met arg thr glu val leu
520/121
GGC TGC CCC GTG ACC CCT GTC TAC AGC TAC TAC GCA CAG AAT GAG GTG GTA ACT ACT GAC
gly cys pro val thr pro val tyr ser tyr trp ala gln asn glu val val thr thr asp
580/141
AGC CTG GAC TTC CGG CAC CAC AGC TAC AAG GAC ATG CGC CAG CTG ATG AAG GCT GTC AAT
ser leu asp phe arg his his ser tyr lys asp met arg gln leu met lys ala val asn
640/161
GAG GAG TGC CCC ACA ATC ACT CGC ACA TAC AGC CTG GGC AAG AGT TCA CGA GGC CTC AAG
glu glu cys pro thr ile thr arg thr tyr ser leu gly lys ser ser arg gly leu lys
700/181
ATC TAC GCA ATG GAA ATC TCA GAC AAC CCT GGG GAT CAT GAA CTG GGC GAG CCC GAG TTC
ile tyr ala met glu ile ser asp asn pro gly asp his glu leu gly glu pro glu phe
760/201
CGC TAC AAG GGC GGC ATC CAC GGC AAT GAG GTG CTA GGC CGA GAG CTC CTG CTC CTC CTC
arg tyr thr ala gly ile his gly asn glu val leu gly arg glu leu leu leu leu
820/221
ATG CAA TAC CTA TGC CAG GAG TAC CGC GAT GGG AAC CCG AGA GTG CGC AAC CTG GTG CAG
met gln tyr leu cys gln glu tyr arg asp gly asn pro arg val arg asn leu val gln
880/241
GAC ACA CGC ATC CAC CTG GTC CCC TCG CTG AAC CCT GAT GGC TAT GAG GTG GCA CGC CAG
asp thr arg ile his leu val pro ser leu asn asn leu asn pro asp gly tyr glu val ala gln
940/261
ATG GGC TCA GAG TTT GGC AAC TGG GCA CTG GGC CTG TGG ACT GAG GAG GGC TTT GAC ATC
met gly ser glu phe gly asn trp ala leu thr arg gly leu trp thr glu glu gly asp ile
1000/281
TTC GAG GAC TTC CCA GAT CTC AAC TCT GTG CTG TGG GCA GCT GAG GAG AAG TGG GTC
phe glu asp phe pro asp leu asn ser val leu trp ala ala glu glu lys lys trp val
1060/301
CCC TAC AGG GTC CCA AAC AAT AAC TTG CCA ATC CCT GAA CGT TAC CTG TCC CAG GAT GCC
pro tyr arg pro pro asn asn asn leu pro ile pro glu arg tyr leu ser pro asp ala
1120/321
ACG GTC TCC ACA GAA GTC CGG GGC ATT ATT CTG TGG ATG GAG AAG AAC CCC TTT GTG CTG
thr val ser thr glu val arg ala ile ile ser trp met glu lys asn pro phe val leu
1180/341
GOT GCA AAT CTG AAC GGT GGT GAG CGC CTT GTG TCT TAT CCC TAT GAC ATG GCC AGC ACA
gly ala asn leu asn gly gly glu arg leu val ser tyr pro tyr asp met ala arg thr
1240/361
CCT AGC CAG GAG CAG CTG TTG GCC GAG GCA CTG GCA GCT GCG CGC GGA GAA GAT GAT GAC
pro ser gln glu gln leu leu ala glu ala leu ala ala arg glu gly asp asp
1300/381
GGG GTG TCT GAG GCC CAG GAG ACT CCA GAT CAC GCT ATT TTC CGC TGG CTG GCC ATC TCA
gly val ser glu ala gln glu thr pro asp his ala ile phe arg trp leu ala ile ser
1360/401
TTT GCC TCC GCC CAT CTC ACC ATG ACG GAG CCG TAC CGG GGA GGC TGC CAG GCC CAG GAC
phe ala ser asp phe ser thr leu his thr asn cys leu glu leu ser val tyr leu gly cys asp
1420/421
TAC ACC AGC GGC ATG GGC ATT GTC AAC GGC GGC AAG TGG AAT CCT CGC TCT GGC ACT TTC
tyr thr ser gly met gly ile val asn gly ala lys trp asn pro arg ser gly thr phe
1480/441
AAT GAC TTT AGC TAC CTG CAC ACA AAT TGT CTG GAG CTC TCC TCA TAC CTG GGC TGT GAC
asn asp phe ser tyr leu his thr asn cys leu glu leu ser val tyr leu gly cys asp
1540/461
AAG TTC CCC CAG GAG AGT GAG CTA CCC CGA GAA TGG GAG AAC AAA GAA GCG CTC CTC
lys phe pro his glu ser glu leu pro arg att arg trp glu asn asn lys glu ala leu leu
1600/481
ACC TTC ATG GAG CAG GTG CAC CGT GGC ATT GAG GGT GTG GTG ACA GAT GAG CAA GGC ATC
thr phe met glu gln val his arg gly ile lys gly val val thr asp glu gln gly ile
1660/501
CCC ATT GCC AAT GCC ACC ATT TCT GTG AGT GGC ATT AAC CAT GGT GTG AAG ACA GCA AGT
pro ile ala asn ala thr ile ser val gly gly ile asn his gly val lys thr ala ser
1720/521
GGA GGT GAC TAC TGG CGC ATT CTG AAC CGG GGT GAG TAC CGT GTG ACA GCT CAC GCA GAG
gly gly asp tyr trp arg ile leu asn pro arg gly glu tyr arg val thr ala his ala glu
1780/541
GCC TAC ACC TCA AGT GCC AAG ATC TGC AAT CTG GAC TAC AAT GGC GCC ACT CAG TGC
gly tyr thr ser ser ala lys ile cys asn val asp tyr asp ile gly ala thr gln cys
1840/561
AAC TTC ATC CTG GCT CGA TCC AAC TGG AAG CCG ATT CGG GAG ATC TTG GCT ATG AAC GGG
asn phe ile leu ala arg ser asn trp lys arg ile arg glu ile leu ala met asn gly
1900/581
AAC CGT CCC ATT CTC CGA GTT GAC CCC TCA CGA CCC ATG ACC CCC CAG CAG CGG CGC ATG
asn arg pro ile leu arg val asp pro ser arg pro met thr pro gln gln **arg arg met**
1960/601
CAG CAG CGC CGT CTA CAG TAC CGG CTC CGC ATG AGG GAA CAG ATG CGA CTG CGT CGC CTC
glu gln **arg arg** glu gln **arg arg** glu **arg arg** glu **arg arg** glu **arg arg** glu **arg arg** glu **arg arg** glu
2020/621
AAT TCT ACC GCA GGC CCT GCC ACA AGC CCC ACT CCT GGC CTT ATG CCT CCC CCT TCC CCT
asn **ser thr ala gly pro ala thr ser pro thr pro ala leu met pro pro pro ser pro**
2080/641
ACA CCA GGC ATT ACC TTG AGG CCC TGG GAA GTT CTA CCC ACT ACC ACT GCA GGC TGG GAG
thr pro ala ile thr leu arg pro thr thr thr thr thr thr thr thr thr thr thr thr thr thr
2140/661
GAG TCA GAG ACT GAG ACC TAT ACA GAA GTA GTG ACA GAG TTT GAG ACA GAG TAT GGC ACT
glu ser glu thr glu thr thr thr glu val val thr glu phe glu thr glu thr glu thr glu thr
2200/681
GAC CTA GAG CTG GAA GAT ALE GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG
asp leu glu val glu glu ile glu glu glu glu glu glu glu glu glu glu glu glu glu glu
2260/701
GGC CTT ACA TTT CCA CTC ACA ACA GTG GAG ACC TAC ACA GTG AAC TTT GGC GAC TTC TGA
gly leu thr phe pro thr thr thr thr val glu thr thr thr val asn phe gly asp phe ---
2320
gactgggatctcaaaagccctgcccaattcaaaactaaggcagcactcccaagcctgtgccagcagcagacatgccatcag
2400
atgtccctgggtggagcccaactcccccagtgaggacatcaaaagctaccgggactctgcagatgactgtgtctaccgcc
2480
ccagctctacctgagccactctggggaggggaggcagggaagcaacgtcacatcaataaagaacgtcatgacaccaa
2560
aaaaaataaaaaaaactcgag
Xho I

b

Disc: 47 **SEAWCS**SVDTM---QYI-VAOCEVPTFMVLAQGRDAD---QW-VTSYKIRYSIDNV 103
A5: 312 **ENKQ**-TPGHEPTV-K---ENLVQVLD-ENLRVBSGOTQOALSKETKKYFVFSKVDISBNGR 372
DDR: 70 **DOAM**-PAGVFPKKEKTLQVLD-ORLMELVALVOTQGRHAGLG-KFYSFSLRYSDROR 132
AEBP1: 12 **DOAM**-AENDES---QTMWIEVDI-RTTTFPGVITQGRDSSIH---DDFVTFVFGSND8Q 70
MFG-88: 331 **ENKMT**-AENDES---ENLVQVLD-OTGRVPTQOARDPQH---IQVFSYKVAISDDOV 389
CF-VIII: 235 **ENKMT**-PQVMDPK---ENLVQVLD-OTGRVPTQOARDPQH---IQVFSYKVAISDDOV 2293
CF-V: 2109 **ENKMT**-AENDES---ENLVQVLD-OTGRVPTQOARDPQH---IQVFSYKVAISDDOV 2167

Disc: 104 **SWPEY**---RMJAA---V---TCGTGRTVVMHFFDTPFIRASIAHPLTW-NHGISLRCEFTYQ 152
A5: 373 **DMITL**---EDGNKHLV-FTGNTDATDVITFPSPKVTIRFVRLRPTWENG-LSLRFELYOC 424
DDR: 133 **RMJON**---KDRWQGV-ISMEDPROVLDKLPVRLVRYFPRADRVMSVCLRVLYOC 185
AEBP1: 71 **TMWY**---TMWY-EDMTFYGVVDKTPVLSLPEFVVARFIRIYPLTWNG-SLCKRLVLYOC 122
MFG-88: 390 **QMTYKRG**-GSR-KV-FQGNLDMHSHKUKIIEKPTMAYRVVLPVWENG-RITLRLLELOC 441
CF-VIII: 2294 **QMTYK**-FQMKY-KV-FQGNQDSPTFVMSLDPPLTRYLRHPQSWH-QIALRMEVLYOC 2345
CF-V: 2168 **ENKMT**YLKSMVDKI-FEGMTNKHVKVNFENPFIISRPVIRVPTWNG-SIALRLLELOC 2221

c

AEBP1: **MQAGAN**KEDDYDGAWCAEDSESQTQWIEVDTRTRTFRFTGVTITQGRDSSIHDDFVTFVFGSND8QTMWY 70
CFE: **NAQRGS**SALLALCALAACNWLQBAQBPAPAGMRERR----- -1
CPN: **MSDLL**SVFLHLLLLPKLVAP----- -1
CPM: **MAISRY**PKANGSGVEVRANDLNDPFCMLGLLLPLVAA----- -1

AEBP1: **TMWY**EDMTFYGVVDKTPVLSLPEFVVARFIRIYPLTWNGSLCKRLVLYOCPTVPSYTAQNEVVTDD 140
CFE: -----LQGED 5

AEBP1: **SILDFR**HSYKMRQLMKAVNEECPTITRTYSLOKSSROGLKIYAMEISINPDHKLGEPEFRYTAG-IGHN 209
CFE: **QISFY**SHRYPRLREALVSWLQCTAISRIYTVGRSFEORLLVLSLWENGVHGPGEPEFRYTAG-IGHN 74
CPN: **VTTFR**SHRYDVLVRLTKVQNECGPTFRVYSIGSRVGRRLVLEPSRHPGHEPLEPEVYV-GRHGN 68
CPM: **LDPMY**HRQEGEAPLKTVAQNYSVTLNLSIGSKVGRNLVNLVVGFRPKKHRIQIFETKYVA-NHGD 68

AEBP1: **KVLGR**ELLMLLQYLQCEYTRDGNFRVRLVQDTRIRHLVPSLNPDOYEVA-QMG-EPG-NWALQMTAE 276
CFE: **KAVGR**ELLFLPLAQYLQCEYTRDGNFRVRLVQDTRIRHLVPSLNPDOYEVA-QMG-EPG-NWALQMTAE 141
CPN: **KALGR**ELMLQSEFLCEFRNRVRIYVQLQDTRIRHLVPSLNPDOYEVA-QMG-EPG-NWALQMTAE 135
CPM: **ETVGR**ELLMLLQYLQCEYTRDGNFRVRLVQDTRIRHLVPSLNPDOYEVA-QMG-EPG-NWALQMTAE 131

AEBP1: **QPDIF**EDPDNLNVLMAAEEKWVYPRVNNKLIPERYLSPDATVSTVETRAIISWENKPPVLNANLNG 346
CFE: **GIDL**NKFPDLDRIYVYNEKSGPH-----NHLKNMKIYDONTKLAPETKAVIRWIMDIPVLSANLNG 207
CPN: **QVDL**NKFPDNLNTIYNEKSGPH-----NHLPLPINWKSQ-----VEPTRAVIRWIMDIPVLSANLNG 197
CPM: **QVDL**NKFPDNLNTIYNEKSGPH-----NHLPLPINWKSQ-----VEPTRAVIRWIMDIPVLSANLNG 175

AEBP1: **GERL**SVSPYDMARTPSBQLAEALAAAGDEDDGVSEAQETPDHAIIFWLAISFASANLWTEPYRGG 416
CFE: **QDLV**ANYPYDETRSGSAH-----EYSSSPDDAIQSLARAYSFNPMASDPNRPFC 258
CPN: **QAVV**ANYPYDKSEFRHVRGVRT-----ASTPTPDNDKLQSLARAYSFNPMASDPNRPFC 251
CPM: **QALV**ASYPYD-----NGVQATQALY-----SRSLPDDVQYQLANTYASRNPMKKKDE-C 225

AEBP1: **QAQDY**TS-QM-GIVNQAQRNFRSOTFMDPSYLTNHC-----LELSVYLOCDKPFHESELPREWENKREALL 480
CFE: **RKND**DDSSVVDOTNGGAMYSVPGQMDPNTLSSNCFEITVELS-----CKFPPEETLKTWEDNONSIL 324
CPN: **G-DY**-----FPDGTQKASYSLSKQMDPNTLSSNCFEITVELS-----CDKFPPEETLQRLWQNRALI 312
CPM: **-KKNQ**N---FPNGTNGTYSWYLLQDQMDYNYIAQCFEITVELS-----CKYPREELKPSFNWNNKASLI 289

AEBP1: **TFMBQ**VIRGIVQVTDQEQIPIANATISVSGINMKVKT---ASGDYWRILNPOEYRVTAHAEGYTSAXI 548
CFE: **SYLE**QIHRGVKGFVRDLQGNFIANATISVSGINDHVDTS---AKDQYWRILNPOEYRVTAHAEGYTSAXI 392
CPN: **QFLR**QWQVQIKWVLDENYWRILANATISVSGINDHVDTS---GHDQYWRILNPOEYRVTAHAEGYTSAXI 380
CPM: **EYIK**QVHLGVQVQVQDQGNFVNVIVQDQKRIHCPIRYNTQYQYVILLPGSYINVTVPDHPHITK 359

AEBP1: **CHVDY**DIGATQCNFILARSNWKIRIILAMKNNPILRVDPSPRMTQQRRMQORRLQYRLNMRQRLR 618
CFE: **VAVP**SPAGVDPELESFSEKSEKSEKELMWNKMSSETINF----- 434
CPN: **VTVQ**PARPTLVNHLKRSIPQVSPVRAPSRHGVRAKVPQARKKEKMRQLQGRPA 438
CPM: **VIT**-PERSQNTSALKDILLFPQOGLSDISVNPSPCMPIPLRYNLDPHSAATKPSLPLFLVLSLHIFPK 427

AEBP1: **RLNST**AGPATSPTPALMPSPSPPTAILTRPKVPLPTTGAWESESETTYTTEVTFETETOTDLEVEIE 688

AEBP1: **KEKEKEKE**KMDTGLTFPLTWTYTYVNFQD 719

repression activity of G4-AEBP1 (summarized in Fig. 4a, lines 3 and 4). In contrast, a GAL4 fusion construct made with the Δ Sty frameshift mutant form of AEBP1 (Fig. 4a, line 2), with CP activity, maintained the repression activity of G4-AEBP1 (Fig. 4c, lane 4). Similarly, all of the AEBP1 mutant derivatives with CP activity retained transcriptional repression activity,

whereas those without CP activity also lacked repression activity (Fig. 4c, d; summarized in Fig. 4a). None of these mutations affect the expression or stability of the fusion proteins (Fig. 4e and data not shown). Thus we have not identified any AEBP1 mutants that retain the transcriptional repression function but fail to exhibit CP activity.

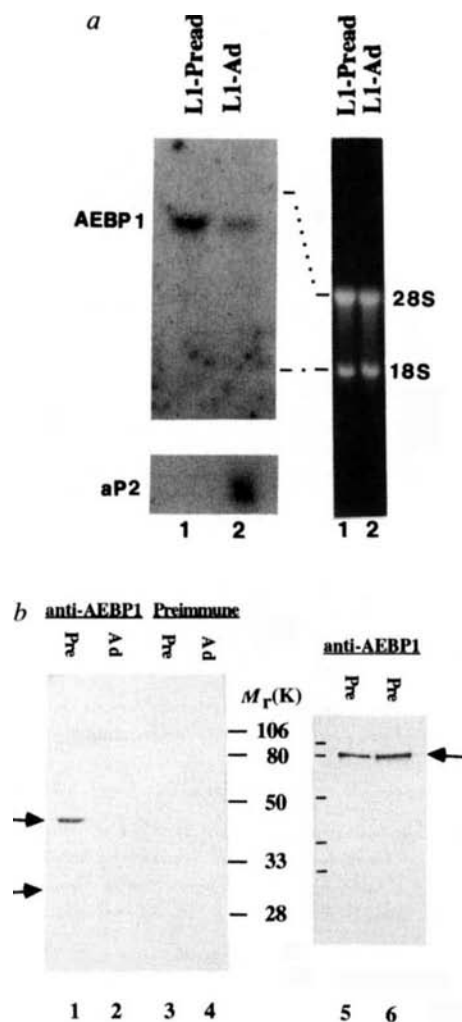


FIG. 2 Downregulation of AEBP1 expression during adipocyte differentiation. *a*, The amount of AEBP1 mRNA in pre- and postdifferentiated adipocytes was analysed by northern blotting. Left, hybridization signals with AEBP1 (top) and aP2 cDNA probes (bottom). Right, ethidium bromide stained gel showing 28S and 18S rRNAs. Lane 1, 3T3-L1 preadipocytes; lane 2, 3T3-L1 adipocytes. *b*, Western blot analysis of total cellular proteins. To determine protein levels during adipocyte differentiation, we first confirmed the deduced amino acid sequence for the full-length AEBP1 protein by *in vitro* translation in a reticulocyte lysate. The expected full-length AEBP1 protein was generated from translation of the cDNA-derived transcript; furthermore, no protein product was detected when a transcript lacking the first two codons was subject to translation (data not shown). These results confirm the deduced amino acid sequence of the ORF (Fig. 1). AEBP1-specific antibodies were raised, which by western analysis detected the *in vitro* translated AEBP1 protein (data not shown). Lanes 1 and 3, 3T3-L1 preadipocytes (Pre); lanes 2 and 4, 3T3-L1 adipocytes (Ad); lanes 5 and 6, two different preadipocyte cell extracts prepared in presence of extra protease inhibitors; lane 1, 2, 5 and 6, anti-AEBP1 serum; lanes 3 and 4, preimmune serum. Arrows indicate the AEBP1-specific protein.

METHODS. Total RNAs (10 μ g) isolated from 3T3-L1 preadipocytes (L1-Pread) and adipocytes (L1-Ad) were analysed with the random-primed 32 P-labelled AEBP1 cDNA as hybridization probe in northern blotting as described²⁸. β -actin hybridization was used to normalize the amount of RNA loaded. The same blot was hybridized with each probe as indicated after stripping. Total cell lysate protein (30 μ g) from 3T3-L1 preadipocytes or adipocytes at 7 days postdifferentiation were resolved by SDS-polyacrylamide (10%) gel electrophoresis and analysed by Amersham's ECL western blotting system with anti-AEBP1 serum. The AEBP1 antibodies were raised in New Zealand white rabbits with the recombinant AEBP1 protein (His-AEBP1; Fig. 4 legend).

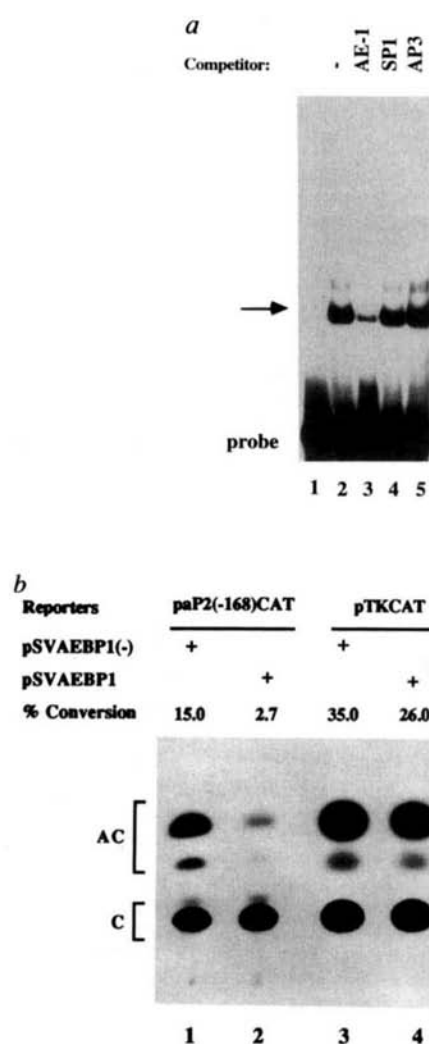
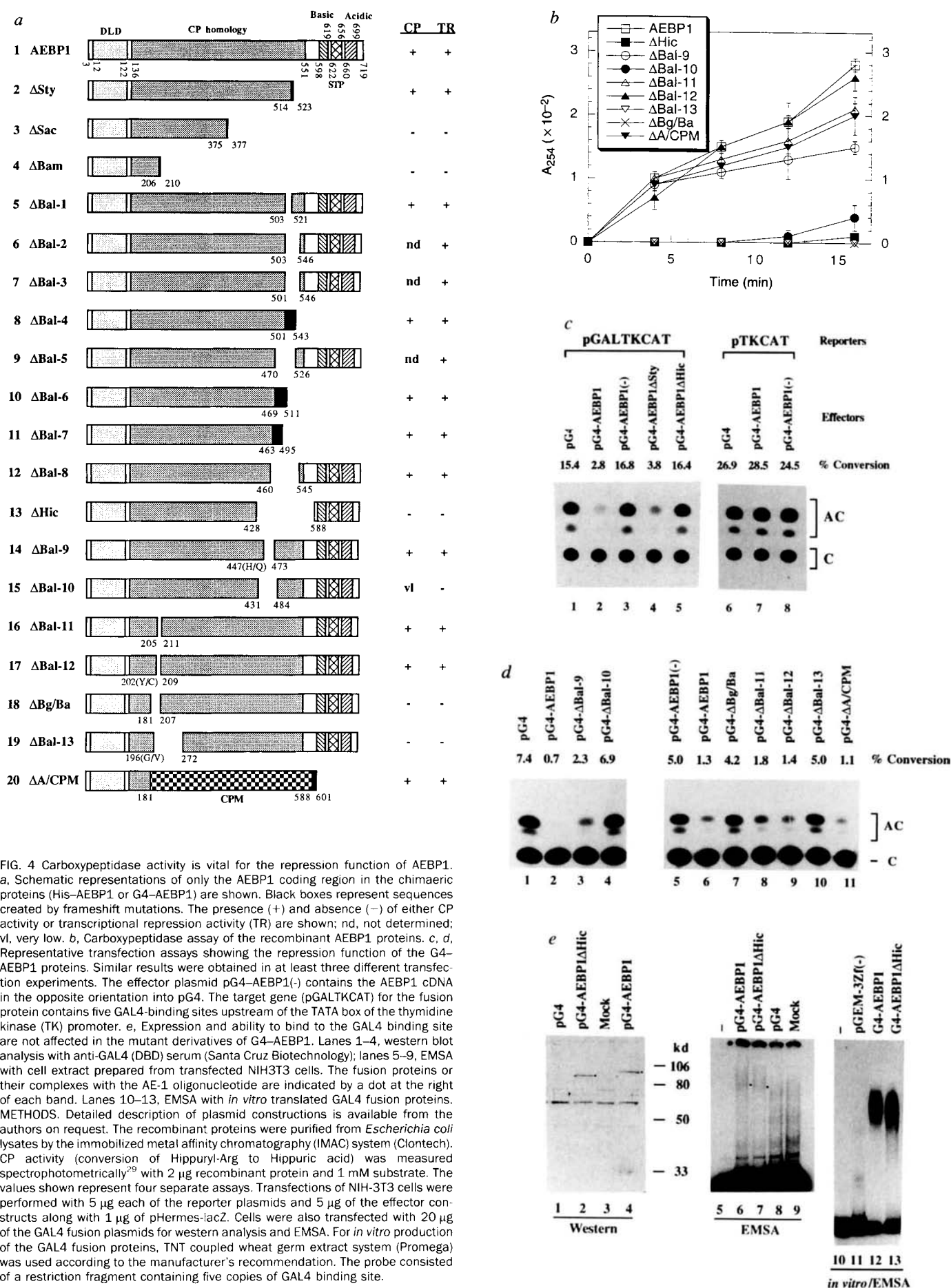


FIG. 3 Functional analysis of AEBP1. *a*, Bacterially synthesized AEBP1 protein binds to the AE-1 oligonucleotide in a sequence-dependent fashion. The histidine-tagged recombinant AEBP1 protein (Fig. 4) was incubated with the AE-1 probe (lane 2) in presence of the indicated competitors (lanes 3–5). Arrow indicates the AEBP1/AE-1 oligonucleotide complex. The slower migrating band might be an AEBP1 dimer/AE-1 complex. Lane 1, probe alone. *b*, AEBP1 is a transcriptional repressor. Transient transfection analysis with the indicated reporter construct along with either the control plasmid pSVAEBP1(-) (lanes 1, 3 and 5) or an expression vector encoding AEBP1 (pSVAEBP1; lanes 2, 4 and 6). The reporter plasmid paP2(-168)CAT (lanes 1 and 2) contains the AE-1 site (sequences -168 to +21 of the aP2 promoter region) inserted upstream of the bacterial CAT gene of plasmid pUC-CAT^{4,5,7}. CAT expression driven by heterologous promoters, such as TK (pTKCAT, see below; lanes 3 and 4), SV40 (pRXV40CAT⁵; lanes 5 and 6) and LTR (pAUCAT⁷; data not shown) sequences, which all lack the AE-1 element, did not alter significantly. The transcription activity is presented in absolute levels of CAT activity (% conversion of chloramphenicol (C) to acetylated chloramphenicol (AC)) driven by each reporter. Similar results were obtained in three different transfection experiments.

METHODS. Recombinant AEBP1 protein (500 ng) was incubated with the 32 P-labelled AE-1 oligonucleotide in the presence of ~200-fold molar excess of the cold unlabelled competitors for electrophoretic mobility-shift assay (EMSA)⁷. The cDNA encoding full-length AEBP1 (from NaeI at nucleotide 149 to XhoI at the 3' end of the cDNA; Fig. 1a) was cloned into the SV40-based expression vector pKSV10 (Pharmacia) in both orientations to construct pSVAEBP1 and pSVAEBP1(-). Transfection of NIH-3T3 cells was performed with 2.5 μ g of each reporter plasmid and 5 μ g of the effector constructs along with 1 μ g of pHermes lacZ, a plasmid that expresses the lacZ gene under the control of the CMV promoter. CAT activity was assayed after adjusting the cell extracts by β -galactosidase activity as described²⁸.



We then investigated whether a heterologous CP domain could replace the CP domain in AEBP1 and still maintain its transcriptional repression function. We made an AEBP1 chimaeric construct that encodes the N terminus of AEBP1 (residues 3–181) fused to the complete sequence, except for the C-terminal 20 amino acids, of the regulatory-type carboxypeptidase M (CPM) (residues 1–407; Fig. 1c). When the GAL4 fusion plasmid encoding this chimaeric protein (pG4- Δ A/CPM, Fig. 4a, line 20) was cotransfected with pGALTKCAT, CAT activity was significantly decreased, as in the case of pG4-AEBP1 (Fig. 4d; compare lanes 6 and 11). No such effect was observed with the control plasmid pG4- Δ A/MPC, in which the CPM coding sequence was inserted in the opposite orientation (data not shown). Furthermore, the histidine-tagged recombinant Δ A/CPM chimaeric protein exhibited CP activity (Fig. 4b). These results convincingly demonstrate that CP activity is important for transcriptional repression function. Thus our findings with mutant derivatives of AEBP1 and the results of the CP domain-swapping experiment strongly suggest that the CP activity of AEBP1 is directly involved in the transcriptional repression mechanism.

Two different mechanisms of transcriptional repression have been proposed. The first involves a passive repression in which the negative regulator either simply blocks the binding of activators or basal factors, or forms inactive heterodimers^{11–15}. The second mechanism involves active repression¹⁶. In this model, a bound repressor counteracts the effects of activators bound at independent sites by direct (inhibitory) interactions with the general transcription factors, the activators themselves, or cofactors (adaptors) required for communication between regulating molecules and the basal machinery for transcription. Our results strongly suggest that AEBP1 is a negative transcription factor with a distinct mechanism involving enzymatic cleavage, by CP activity, of proteins involved in transcription. Targeting a protease to a promoter in transcriptional repression mechanism would represent a new feature in transcription regulation. □

Received 3 August; accepted 11 September 1995.

1. Spiegelman, B. M., Frank, M. & Green, H. J. *biol. Chem.* **258**, 10083–10089 (1983).
2. Bernlohr, D. A., Angus, C. W., Lane, M. D., Bolanowski, M. A. & Kelly, T. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5468–5472 (1984).
3. Chapman, A. B., Knight, D. M., Dieckman, B. S. & Ringold, G. M. *J. biol. Chem.* **259**, 15548–15555 (1984).
4. Distel, R. J., Ro, H.-S., Rosen, B. S., Groves, D. L. & Spiegelman, B. M. *Cell* **49**, 835–844 (1987).
5. Herrera, R., Ro, H.-S., Robinson, G. S., Xanthopoulos, K. G. & Spiegelman, B. M. *Molec. cell. Biol.* **9**, 5331–5339 (1989).
6. Christy, R. J. *et al. Genes Dev.* **3**, 1323–1335 (1989).
7. Ro, H.-S. & Roncari, D. A. K. *Molec. cell. Biol.* **11**, 2303–2306 (1991).
8. Yang, V. W., Christy, R. J., Cook, J. S., Kelly, T. J. & Lane, M. D. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3629–3633 (1989).
9. Ma, J. & Ptashne, M. *Cell* **50**, 137–142 (1987).
10. Shi, Y., Seto, E., Chang, L.-S. & Shenk, T. *Cell* **67**, 377–388 (1991).
11. Akerblom, I. E., Slater, E. P., Beato, M., Baxter, J. D. & Mellon, P. L. *Science* **241**, 350–353 (1988).
12. Graupner, G., Willis, K. N., Tzukerman, M., Zhang, X.-K. & Pfahl, M. *Nature* **340**, 653–656 (1989).
13. Jonat, C. *et al. Cell* **62**, 1189–1204 (1990).
14. Schule, R. *et al. Cell* **62**, 1217–1226 (1990).
15. Yang-Yen, H.-F. *et al. Cell* **62**, 1205–1215 (1990).
16. Renkawitz, R. *Trends Genet.* **6**, 192–197 (1990).
17. Hoffmann, A. *et al. Nature* **346**, 387–390 (1990).
18. Poole, S., Firtel, R. A., Lamar, E. & Rowekamp, W. J. *molec. Biol.* **153**, 273–289 (1981).
19. Takagi, S. *et al. Neuron* **7**, 295–307 (1991).
20. Johnson, J. D., Edman, J. C. & Rutter, W. J. *Proc. natn. Acad. Sci. U.S.A.* **90**, 5677–5681 (1993).
21. Stubbs, J. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 8417–8421 (1990).
22. Kane, W. H. & Davie, E. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6800–6804 (1986).
23. Wood, W. I. *et al. Nature* **312**, 330–337 (1984).
24. Manser, E. *et al. Biochem. J.* **267**, 517–525 (1990).
25. Gebhard, W., Schube, M. & Eulitz, M. *Eur. J. Biochem.* **178**, 603–607 (1989).
26. Tan, F., Chan, S. J., Steiner, D. F., Schilling, J. W. & Skidgel, R. A. *J. biol. Chem.* **264**, 13165–13170 (1989).
27. Singh, H. *et al. Cell* **52**, 415–423 (1988).
28. Xu, L. *et al. Nucleic Acids Res.* **22**, 646–655 (1994).
29. Folk, J. E. *et al. J. biol. Chem.* **235**, 2272–2277 (1960).

ACKNOWLEDGEMENTS. G.-P.H. and A.M. contributed equally to this work. We thank R. Singer and G. Johnston for discussion and comments on the manuscript; T. Shenk for the plasmids pG4 and pGALTKCAT; R. A. Skidgel for the human CPM cDNA plasmid; M. E. Fotaki for the plasmid pHermes-lacZ; and M.-L. Zhu, Y. Zhang and A. Young for technical assistance. This work was supported by grants from the Heart and Stroke Foundation (Nova Scotia) of Canada and NSERC to H.-S.R.; H.-S.R. is a scholar of the HSFC.

N-glycans as apical sorting signals in epithelial cells

Peter Scheiffele, Johan Peränen & Kai Simons*

Cell Biology Programme, European Molecular Biology Laboratory, Postfach 102229, 69112 Heidelberg, Germany

IN epithelial Madin–Darby canine kidney (MDCK) cells newly synthesized molecules are sorted in the *trans*-Golgi network and directly delivered to their apical and basolateral surface destinations¹. Sorting is mediated by signals in the cytoplasmic domains of basolateral transmembrane proteins^{2–4} whereas glycosylphosphatidylinositol-linked proteins have apical sorting information in their glycolipid tails^{5,6}. Signals for apical transmembrane proteins are thought to reside in their ectodomains, because truncated forms lacking the cytoplasmic tail and the membrane anchor are secreted apically^{7,8}. Here we demonstrate that carbohydrates act as an apical targeting signal for secretory proteins. Growth hormone, which is non-glycosylated and secreted from both sides of MDCK cell layers⁹, is secreted from the apical side when glycosylated. Thus glycans not only play a general role in protein folding¹⁰ but also appear to function in protein sorting in biosynthetic traffic.

A role for N-glycans in apical and basolateral protein sorting of membrane proteins was excluded by early studies using tunicamycin treatment, which led to accumulation of most of the newly synthesized protein in the endoplasmic reticulum¹¹. The small fraction that was delivered to the cell surface was analysed by qualitative methods. Later it was found, however, that the major apically secreted glycoprotein in MDCK cells, clusterin (gp80), is secreted randomly to both sides after tunicamycin treatment¹². To re-explore this question we have now analysed the secretion of a non-glycosylated protein and compared its secretion with mutants of the same protein that carry engineered N-glycosylation sites. We have used the non-glycosylated rat growth hormone (GH) which when expressed in MDCK cells is known to be secreted to both the apical and the basolateral sides⁹. Guan *et al.* introduced one (GH2) or two (GH12) N-glycosylation sites into this protein and these mutant forms were efficiently secreted from fibroblasts¹³.

After transfecting MDCK cells with the wild-type and mutant cDNAs, stably expressing clones were selected and analysed. Secreted proteins were collected and equivalent amounts of the media and cell lysates were subjected to western blotting (Fig. 1). The observed bands of lower mobility are N-glycosylated forms as demonstrated by their sensitivity to N-glycosidase F (Fig. 1, lanes 8–13). About 75% of the mutant proteins were glycosylated. Less than 10% of the protein was detected inside the cells, demonstrating efficient secretion (Fig. 1, lanes 2–7).

To ensure that the polarized character of the selected cell clones was preserved, the distribution of the 114K apical and 58K basolateral marker proteins was determined¹⁴. Indirect immunofluorescence using monoclonal antibodies 4.6.5a and 6.23 showed both proteins to be localized as in the parental cell line (data not shown). Second, proteins secreted apically and basolaterally radiolabelled with ³⁵S-methionine were analysed by two-dimensional gel electrophoresis. The patterns of the endogenous proteins observed were identical to those of untransfected cells (data not shown). Therefore we conclude that the obtained cell clones are polar and have retained the ability to target proteins correctly to the apical and the basolateral sides.

Fully polarized cells grown on polycarbonate filters were used to test the polarity of secretion of wild-type and mutant proteins (Fig. 2). The unglycosylated GH was secreted 34 ± 8% (n = 3) into the apical medium and 66 ± 8% (n = 3) into the

* To whom correspondence should be addressed.

basolateral. This distribution is in accordance with results described previously for other unglycosylated proteins when expressed in MDCK cells^{8,9,15}. In contrast, the mutant protein GH2 carrying one carbohydrate chain at position Asn 95 was clearly enriched in the apical medium ($66 \pm 2\%$, $n=3$). Most importantly, the doubly glycosylated GH12 carrying two *N*-glycans at position 19 and 95 was almost exclusively ($92 \pm 1\%$, $n=3$) transported to the apical compartment. When *N*-glycosylation of GH12 was blocked using tunicamycin, secretion was randomized (lanes 7, 8). As an internal control we followed the secretion of the endogenous apical marker protein gp80, which in all clones showed about the same distribution (75% apical and 25% basolateral) as had been reported previously¹². A time course of the secretion showed a constant ratio of apically and basolaterally recovered GH or GH12 at time points later than 1 h (Fig. 3). The lag in basolateral secretion observed for the 30 min time point is due to the time required for diffusion through the filter¹⁵.

In the cell clones expressing the glycosylated proteins we also found that the unglycosylated fraction was enriched in the apical medium. The formation of mixed dimers or multimers between glycosylated and unglycosylated GH probably accounts for the apical sorting of the unglycosylated form: GH dimers are present in secretory granules, and are induced by the chemical conditions in the lumen of the *trans*-Golgi network¹⁶. Multimerization has been proposed to be a general mechanism for the segregation of molecules for regulated secretion from the bulk of constitutively secreted proteins¹⁷. Similar conditions may prevail in the *trans*-Golgi network of MDCK cells.

The conclusion from these experiments is that the introduction of *N*-glycosylation sites into rat GH leads to apical sorting and delivery of this secretory protein in MDCK cells. Interestingly, the introduction of two glycosylation sites led to increased apical delivery as compared to monoglycosylation. The generality of this conclusion is supported by a surprising number of already published studies, none of which have inferred that *N*-glycans play a role as apical sorting signals (Table 1). All proteins reported to be apically secreted contain *N*-glycans (with one exception, the hepatitis B surface antigen¹⁸ discussed in ref. 19). Moreover, in three cases the presence of the *N*-glycans was shown to be critically involved in apical delivery^{12,20,21}. On the other hand, non-glycosylated secretory proteins are usually secreted both apically and basolaterally^{8,9,15,22}, presumably being included by default into the post-Golgi transport vesicles. We therefore conclude that *N*-glycans indeed function as a general sorting signal for apical secretory proteins in MDCK cells.

Some light on the nature of the oligosaccharide site that could be involved in apical sorting comes from studies using either

TABLE 1 Polarity of secretion of cellular proteins from MDCK cells

Apical		Non-polarized	
Clusterin (gp80) ¹²	<i>N</i> -glycosylation required	Growth hormone ⁹	No <i>N</i> -glycosylation present
Erythropoietin ²¹		Prochymosin ⁹	
hCBG ²⁰		α -2 μ -globulin ⁹	
NEP (sec) ⁷	<i>N</i> -glycosylation present	Lysozyme ^{9,15}	
APN (sec) ⁸		Uteroglobin ²²	
plg-R (sec) ³		Cystatin C ⁸	
DAF (sec) ⁶		uPA (mouse) ²⁷	
PLAP (sec) ⁵			

When polarized secretion was lost upon removal of *N*-glycans by mutagenesis or tunicamycin treatment, examples were listed as '*N*-glycosylation required'. 'Sec' indicates anchor-minus forms of membrane proteins which were expressed as soluble secreted proteins. Secretory forms of viral proteins have been excluded from the list because of conflicting results.

hCBG, Human corticosteroid binding globulin; NEP, neutral endopeptidase; APN, aminopeptidase N; plg-R, polymeric immunoglobulin receptor; DAF, decay-accelerating factor; PLAP, placental alkaline phosphatase; uPA, urokinase-type plasminogen activator.

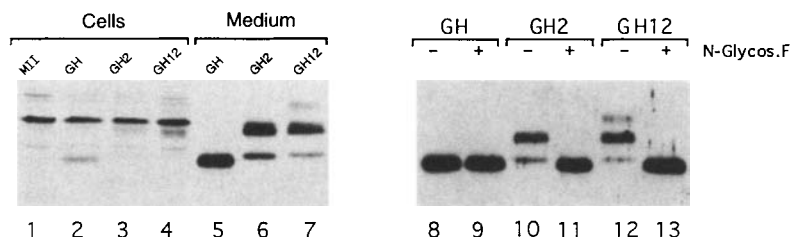
mutant MDCK cell lines defective in terminal glycosylation or inhibitors of oligosaccharide processing. None of the modifications that affect the *N*-glycan side chains after exit from the endoplasmic reticulum seem to play any role in apical sorting^{23,24}. Thus the putative lectin that is responsible for recognizing the *N*-glycans in the *trans*-Golgi network probably binds to the high-mannose core of the side chains. One lectin candidate has been identified in apical transport vesicles, VIP36²⁵. However, the binding specificity of this potential lectin is unknown.

If *N*-glycans are used as apical sorting signals, the *N*-glycosylated basolateral secretory proteins must somehow escape from being included into the apical transport vesicles. One obvious possibility is that these proteins are sorted by receptors binding to the secretory proteins with higher affinity than the putative apical lectin. The sorting efficiency could also depend on the number of sterically available glycan chains. In keeping with such a selective handling of secretory proteins in the *trans*-Golgi network, treatment with NH₄Cl randomizes the secretion of basolaterally secreted proteins whereas this acidotropic agent has no effect on apically sorted secretory proteins^{20,26,27}. We also observed that NH₄Cl had no effect on the apical secretion of GH12 (data not shown).

It remains to be determined whether membrane proteins also use glycans as sorting signals. Apical transmembrane glycoproteins are known to carry apical signals in their ectodomains⁴, consistent with a role for glycosylation in the sorting process. Glycosylation could provide a clue for understanding the perplexing plasticity of protein sorting in epithelial cells. There are

FIG. 1 Western blot analysis of media and cell lysates of MDCK cell clones expressing unglycosylated (GH) and glycosylated forms (GH2, GH12) of rat growth hormone probed with an anti-GH antibody. *N*-glycosylation of the products with lower mobility is confirmed by digestion with *N*-glycosidase F (lanes 8–13). The additionally detected proteins in the cell lysates (lanes 1–4) are crossreacting endogenous proteins because they are also present in lysates of the parental MDCK cells (MII).

METHODS. Stably transfected MDCKII cells were grown on 35-mm dishes to 80% confluency, washed three times with phosphate buffered saline (PBS) and incubated in DMEM at 37 °C, 5% CO₂ for 3 h. The medium was taken off, cells washed in PBS and lysed in 2% NonidetP-40, 0.2% SDS, 1 mM DTT in PBS on ice. Proteins were precipitated from lysate and medium with a final concentration of 10% trichloroacetic acid. When indicated, samples were treated over night with *N*-glycosidase F (Boehringer) according to the manufacturer's instructions. Equivalent amounts were analysed on 12.5% SDS polyacrylamide gels. Proteins were detected by western blotting with rat growth hormone rabbit polyclonal antiserum (Chemicon) followed by anti-rabbit horseradish peroxidase-conjugated antibodies



and ECL (Amersham). MDCKII cells were transfected with wild-type and mutant constructs using the retroviral system³⁰. The Kozak sequence in pgGH, pgGH2 and pgGH12 (provided by J. K. Rose, described in ref. 13) was changed to 5'-... ACC ATG G...-3' by PCR and constructs were verified by sequencing. The gene was placed under control of the LTR promoter in expression vector pLXSN and transfection was as described³⁰. Cells were cultured as described¹ and stably expressing clones were selected with 0.5 mg ml⁻¹ G418.

FIG. 2 Glycosylated growth hormone forms are secreted polarized from filter-grown cells as analysed by western blotting using anti-GH (α GH) and anti-gp80 (α gp80) antibodies. When glycosylation in GH12 is prevented by tunicamycin (TUN, lanes 7, 8) secretion is randomized.

METHODS. Cells were seeded on 24-mm polycarbonate filters (Costar) and grown for 3.5 days. Filters were washed three times in PBS containing 0.5 mM $MgCl_2$ and 0.9 mM $CaCl_2$ and incubated in DMEM at 37 °C, 5% CO_2 . Secreted proteins were collected for 3 h or indicated times, recovered from apical (AP) or basolateral (BL) culture medium and analysed by western blotting (a). The same nitrocellulose membrane was then stripped and reprobed with a polyclonal gp80 antiserum (provided by C. Koch-Brandt). Results were quantified by densitometric scanning and averaged from three independent experiments (b). Similar results were obtained using three different clones of each cell line. When desired, cells were preincubated in DMEM containing $12 \mu g\ ml^{-1}$ tunicamycin (Sigma) for 1.5 h, washed and then fresh medium containing tunicamycin was added. For treatment with NH_4Cl , preincubation was for 30 min in 10 mM NH_4Cl in DMEM containing 20 mM HEPES pH 7.4. For two-dimensional polyacrylamide electrophoresis (2D-PAGE) cells grown on filters were continuously labelled with $300 \mu Ci\ ^{35}S$ -methionine for 3 h. Proteins were recovered from the media as described above and equivalent amounts were analysed by 2D-PAGE and autoradiographed. For indirect immunofluorescence, cells were fixed using the pH-shift paraformaldehyde method. Apical and basolateral marker antigens were stained with monoclonal antibodies 4.6.5a or 6.23 (ref. 14), respectively, followed by an anti-mouse IgG-FITC (Sigma) conjugated antibody. Confocal laser scan microscopy was performed with the EMBL compact confocal system.

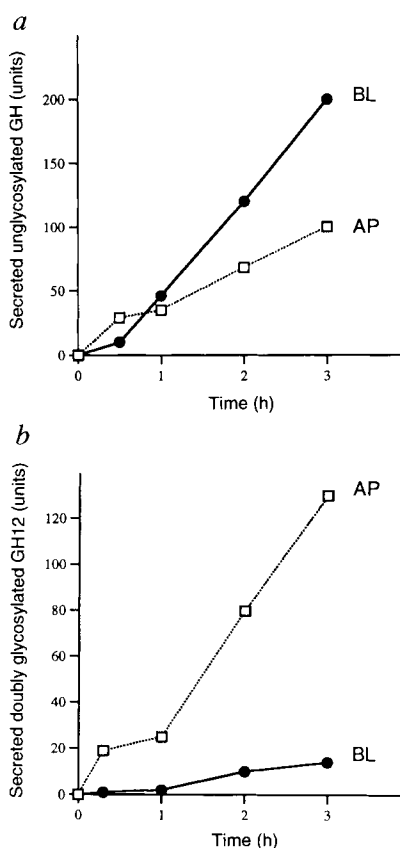
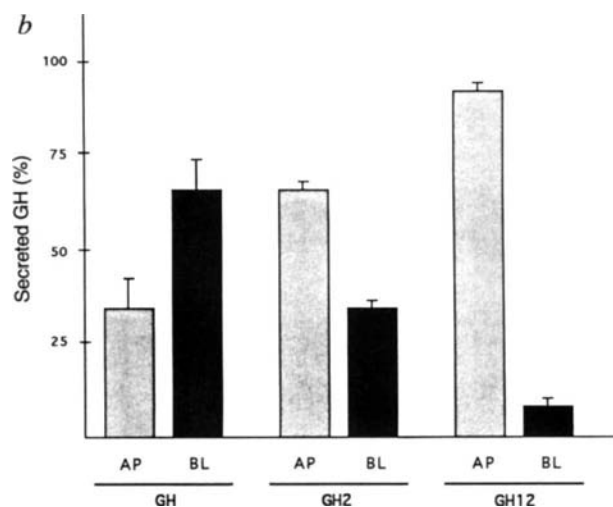
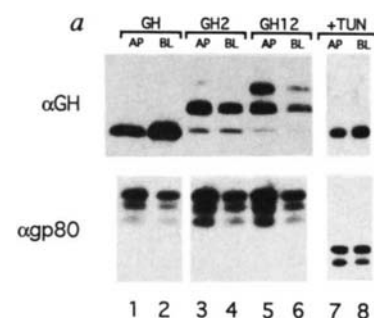


FIG. 3 Time course of secretion of unglycosylated GH (a) and doubly glycosylated GH12 (b) into the apical (□) and basolateral (●) media. Secreted proteins were analysed as described in Fig. 2. Amounts are given in arbitrary units from densitometric scanning of western blots with anti-GH antibodies.

promiscuous luminal signal to come into play and redirect the protein to the apical side. □

Received 3 July; accepted 21 September 1995.

- Wandinger-Ness, A., Bennet, M. K., Antony, C. & Simons, K. *J. Cell Biol.* **111**, 987–1000 (1990).
- Hunziker, W., Harter, C., Matter, K. & Mellman, I. *Cell* **66**, 1–20 (1991).
- Mostov, K. E., Harris, J. M. & Breitfeld, P. P. *J. Cell Biol.* **105**, 2031–2036 (1987).
- Matter, K. & Mellman, I. *Curr. Opin. Cell Biol.* **6**, 545–554 (1994).
- Brown, D. A., Crise, B. & Rose, J. K. *Science* **245**, 1499–1501 (1989).
- Lisanti, M. P., Caras, I. W., Davitz, M. A. & Rodriguez-Boulant, E. *J. Cell Biol.* **109**, 2145–2156 (1989).
- Corbeil, D., Boileau, G., Lemay, G. & Crine, P. *J. Biol. Chem.* **267**, 2798–2801 (1992).
- Vogel, L. K., Spiess, M., Sjöström, H. & Norén, O. *J. Biol. Chem.* **267**, 2794–2797 (1992).
- Gottlieb, T. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2100–2104 (1986).
- Helenius, A. *Molec. Biol. Cell* **5**, 253–265 (1994).
- Green, R. F., Meiss, H. K. & Rodriguez-Boulant, E. *J. J. Cell Biol.* **89**, 230–239 (1981).
- Urban, J., Parczyk, K., Leutz, A., Kayne, M. & Kondor-Koch, C. *J. Cell Biol.* **105**, 2735–2743 (1987).
- Guan, J.-L., Machamer, C. E. & Rose, J. K. *Cell* **42**, 489–496 (1985).
- Balcarova-Ständer, J., Pfeiffer, S. E., Fuller, S. D. & Simons, K. *EMBO J.* **3**, 2687–2694 (1984).
- Kondor-Koch, C. B., Bravo, R., Fuller, S. D., Cutler, D. & Garoff, H. *Cell* **43**, 297–306 (1985).
- Cunningham, B. C., Mulkerrin, M. G. & Wells, J. A. *Science* **253**, 545–548 (1991).
- Chanat, E. & Huttner, W. B. *J. Cell Biol.* **115**, 1505–1519 (1991).
- Gonzalez, A., Nicovani, S. & Juica, F. *J. Biol. Chem.* **268**, 6662–6667 (1993).
- Fiedler, K. & Simons, K. *Cell* **81**, 1–20 (1995).
- Musto, N. A. *Exp. Cell Res.* **209**, 271–276 (1993).
- Kitagawa, Y. *et al. Exp. Cell Res.* **213**, 449–457 (1994).
- Vogel, L. K., Suske, G., Beato, M., Norén, O. & Sjöström, H. *FEBS Lett.* **330**, 293–296 (1993).
- Meiss, H. K., Green, R. F. & Rodriguez-Boulant, E. *Molec. cell. Biol.* **2**, 1287–1294 (1982).
- Wagner, M., Morgans, C. & Koch-Brandt, C. *Eur. J. Cell Biol.* **67**, 84–88 (1995).
- Fiedler, K., Parton, R. G., Kellner, R., Etzold, T. & Simons, K. *EMBO J.* **13**, 1729–1740 (1994).
- Caplan, M. J. *et al. Nature* **329**, 632–635 (1987).
- Ragno, P. *et al. Exp. Cell Res.* **203**, 236–243 (1992).
- Rodriguez-Boulant, E. & Powell, S. K. *A. Rev. Cell Biol.* **8**, 395–427 (1992).
- Casanova, J. E., Breitfeld, P. P., Ross, S. A. & Mostov, K. E. *Science* **248**, 742–745 (1990).
- Miller, A. D. & Rosman, G. M. *Biotechniques* **7**, 980–990 (1989).

ACKNOWLEDGEMENTS. We thank J. Rose for the growth hormone cDNA constructs; C. Koch-Brandt for the gp80 antibodies; H. Virta for cell culture; S. Ruf for technical assistance; S. Eaton and M. Zerial for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft and the HCM programme.

many examples of membrane proteins that are basolateral in one cell type and apical in another²⁸. It is also known that, when the basolateral signal is removed by mutagenesis from the cytoplasmic domain of a basolateral glycoprotein, the mutated protein will be delivered to the apical membrane by virtue of its ectodomain signal⁴. Basolateral proteins, therefore, seem to contain both a basolateral and an apical signal but the former is dominant over the other. Modulation of protein sorting could simply be brought about by inhibition of the cytoplasmic signal by post-translational modifications²⁹ which would allow the

Illegitimate concerns

Bryan Sykes

The Quest for Anastasia: Solving the Mystery of the Lost Romanovs. By John Klier and Helen Mingay. *Smith Gryphon*: 1995. Pp. 246. £15.99.

Queen Victoria's Gene: Haemophilia and the Royal Family. By D. M. and W. T. W. Potts. *Alan Sutton*: 1995. Pp. 160. £16.99.

IN the early hours of Wednesday, 17 July 1918, the Russian royal family was taken to the basement of a house in a grim industrial city in the Urals and shot. Their bodies were dumped in the back of a lorry, taken to a nearby wood and thrown down a mine-shaft. Two days later they were pulled out and hastily buried in a shallow grave. There they remained until 1979 when, after years of searching, a local geologist found the grave. Naturally reluctant to publicize the discovery at a time when the Communist Party still had an iron grip on the Soviet Union, he reburied what he had found. (Only two years earlier, President Leonid Brezhnev had in fact instructed the then regional First Secretary, Boris Yeltsin, to flatten the house in which the Romanovs had been held for fear of its becoming a focus for a monarchist revival.) The bones had to wait another 12 years to see the light of day again, by which time the political climate had changed. Although there was never much question about the authenticity of the bones, what doubts remained were silenced when DNA was recovered and matched with living relatives.

Both these books take their cue from the publication of these DNA results to update the story of the Romanovs and, in particular, to explore the connections with the only two skeletons that were never discovered — the 17-year-old Anastasia and her younger brother the Tsarevich Alexei. In *The Quest for Anastasia*, John Klier, a professor of Russian history, and Helen Mingay, a journalist, provide a thorough and readable account of the career of the most famous, but by no means the only, pretender. Dragged freezing from a Berlin canal in 1920, the unknown woman "Fraulein Unbekannt", or Anna Anderson as she became known, never doubted that she was the Grand Duchess Anastasia. Despite some glaring anomalies (she spoke no Russian, for instance), she spent her life alternating between mental asylums and the protection of utterly devoted supporters who shared her conviction.

Equally, her claims were vigorously rejected by her closest 'relatives' after meetings that illustrated her evident eccentricities. When her putative aunt, Princess Irene of Prussia, was eventually persuaded to meet her, Anna ran away from the dinner-table and hid in her

room. Hardly the behaviour of a ruthless gold-digger eager to get her hands on what was left of the Romanov fortune. This uncertainty, this polarity of opinion, lasted for the rest of her life (she died in 1984 aged 83). After the romance of claim and counter-claim, the definitive DNA testing seems positively prosaic.



The daughters of Tsar Nicholas II: Olga, Tatiana, Marie and Anastasia.

Anna Anderson was not Grand Duchess Anastasia and that's that.

The authors finish their gripping tale by guiding us through the legal and political bickering surrounding access to the biopsy sample that was to provide the ultimate source of DNA (a process ironically, and expensively, opposed by RNA — the Russian Nobility Association in this instance). Whoever she was, Anna Anderson was a remarkable and hypnotic woman but, unfortunately, the book throws little light on how she could have known the minute details of life at the Imperial Court that so thoroughly convinced her disciples.

By contrast, *Queen Victoria's Gene* moves backwards from 1918 and concentrates not on Anastasia but on Tsarevich Alexei, her younger brother. Although his body was also missing from the grave, in the absence of a convincing pretender, no one doubts he was murdered in 1918 with the rest of the royal family. Alexei suffered from haemophilia, which he inherited from Queen Victoria of England. It is this gene that provides the thread for what is a popular history of the interbreeding European monarchy over two centuries. The authors paint a picture of greed, lust and debauchery occasionally relieved by devotion (as in

Victoria and Albert) that serves as a backdrop to their most outrageous suggestion — that Queen Victoria was probably illegitimate.

There is no dispute that Victoria was a symptomless carrier of the haemophilia gene. (The gene is carried on the X chromosome, so females with two X chromosomes can mask the haemophilia allele on one of them with its normal counterpart; men, with only one X, aren't so lucky.) But did Victoria inherit the gene or did it arise as a spontaneous mutation? Adding what was already known to their own new research, the authors, a doctor and a biologist, find no evidence of haemophilia in her recorded ancestors, so they conclude either that the defective gene was a new mutation, as it

is in roughly a third of cases, or that she inherited it from an unknown and unidentified lover of her mother, Victoire.

So far, so good. Then, in an astonishing misuse of probability statistics, they attempt to persuade us that the second explanation is the more likely. Not only would the lover have had to be an adult haemophiliac — unlikely but not impossible even in the nineteenth century — but he would also have had to be quick. Victoire arrived from Germany to marry the Duke of Kent in May 1818 and Victoria was born almost exactly a year later. Unable to identify the nimble little bleeder, the authors attempt to pass off a perversion of probability theory on their unwary readership. "Victoire might have slept with another man... but would she have chosen a haemophiliac? The chances are small but substantially greater than the chance of mutation." *Ergo* Victoria was more likely to be illegitimate than not. This is, to say the least, a naive analysis of the odds because the authors neglect the large prior expectation of legitimacy. The fact that Victoria was a carrier of haemophilia has virtually no bearing on whether or not she was illegitimate.

Elsewhere the book pushes the theory

From Queen Victoria's Gene

of historical contingency, as Stephen Jay Gould calls it (chance to the rest of us), to preposterous limits. If only, they suggest, Victoria's haemophilic son Leopold had fallen down the stairs and died a few months earlier, then his son, Charles Edward, would not have been born. Hitler would then have lost an influential supporter and 30 million people would have been spared. They even suggest this happy outcome if Charles Edward had inherited his father's lethal gene. Wrong again. Sons inherit their single X chromosome from their mothers, so it would have been quite impossible for Leopold to have passed the haemophilia gene to his son. The only remotely persuasive historical consequence of the gene was the part it played in the rise of Rasputin to a position of influence in the Russian royal family. His hypnotic powers, so we are told, could alleviate the painful symptoms of the young Tsarevich Alexei, something the court physicians had failed to do.

As in *The Quest for Anastasia*, there is very little genetics in this book and even what there is is tinged with the depressingly negative view of humanity in general and monarchies in particular that drips from the pages. Even the five lines devoted to Watson and Crick's discovery of the structure of DNA wastes one of them by telling us they worked "in temporary accommodation".

It is a pity that the authors' account of their detailed work on this most famous of genes has been spoiled by carelessness, inaccuracies and distortions (the UK Forensic Science Service, for example, is based in Aldermaston, not Amersham). Although these infelicities may go unnoticed by the intended general readership, they do nothing to promote a real appreciation of genetics. □

Bryan Sykes is at the Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

Putting the peas in context

Peter J. Bowler

Gregor Mendel: The First Geneticist. By Vitezslav Orel. Oxford University Press: 1995. Pp. 353. £25.

MENDEL may be unique in the annals of modern science. According to conventional wisdom, he discovered the principles of a new science yet died before those principles were accepted. Mendel's failure to gain recognition in his own lifetime seems to enhance the romantic attraction of his story. Who can fail to be moved by the image of the painstaking experiments in the monastery garden that were ignored by all contemporaries?

Vitezslav Orel is head of the Mendelianum at the Moravian Museum in Brno, Austria, established in the monastery where Mendel worked, and is thus at the nerve-centre of international Mendel research. He is in an unrivalled position to provide a detailed account of Mendel's life and career. His biography is clearly destined to become the standard source, replacing old favourites such as Hugo Iltis's 1924 account. Orel is particularly concerned to trace Mendel's education, showing that he was by no means the untutored amateur of popular myth. On the contrary,

his scientific education laid the foundations for his unique experiments with peas, in which he traced the transmission of hereditary characters. Orel also provides detailed insights into Mendel's other activities: his research on bees and on meteorology (for which he did gain some recognition) and his efforts as abbot of his monastery. Here we see the character of the man: efficient, enthusiastic and at a crucial stage willing to strike out into a new method of studying nature.

There is, of course, a detailed chapter outlining Mendel's classic experiments, published in 1866. Considering the brevity of his paper, and the relative paucity of other written material left by Mendel, historians have come up with a surprising number of different interpretations of his intentions. Orel successfully out-manoeuvres both those who would turn Mendel into a pure empiricist and those who have suggested that he cheated in order to substantiate his preconceived hypothesis.

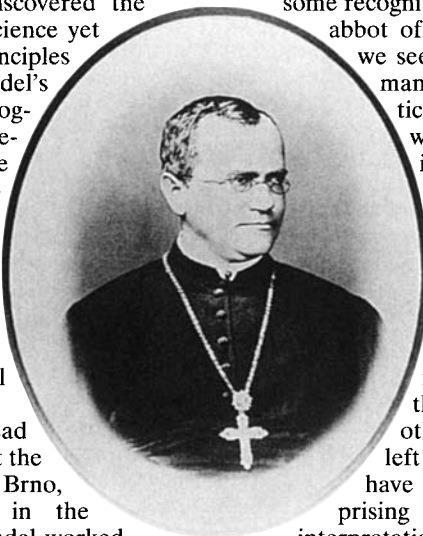
He is less effective, though, in dealing with the challenge offered by historians who argue that the story of Mendel's discovery is a myth created by the founders of genetics in the early twentieth century. It has been suggested that a careful

reading of the 1866 paper shows that Mendel did not in fact propose the modern concept of paired characters linked to factors (genes) in the reproductive cells. In effect, the modern ideas were read into Mendel's paper by the rediscoverers (Karl Correns and Hugo de Vries in 1900, and later William Bateson). On this interpretation, however prescient the experimental demonstrations of character-transmission, Mendel's own real interest was the formation of hybrids, not the laws of heredity.

Orel certainly warns his readers that there have been such iconoclastic claims, but, as his subtitle indicates, his account is firmly within the orthodox tradition. His discussion of "Mendelian myth-making" comes after his analysis of the experiments, an analysis that has already begged the question by translating the results into modern terms, diagrams and all. At one point Orel dismisses the revisionists as "hero-bashers". But historians of genetics as eminent as Robert Olby have argued that Mendel was not a Mendelian in the sense defined by his posthumous followers. Such warnings cannot be set aside quite so easily; we must acknowledge at least the possibility that the context of Mendel's own work differed substantially from that of the rediscoverers. Placing Mendel's work in its context means more than making a sympathetic study of his career; it requires a willingness to uncouple his key text from the layers of meaning read into it by later geneticists.

Orel's discussion of events following Mendel's death concentrates as much on the celebrations of his work as it does on the actual 'rediscovery'. One could hardly expect a biography of Mendel to spend too much time on later developments, but the question of how Mendel's own work was transformed by the rediscoverers is more fundamental. The iconoclasts are not suggesting that his work was without influence in 1900: whatever Mendel's own intentions, his demonstrations forced Correns, De Vries and Bateson to rethink their whole approach to heredity. It would become even more important to see Mendel's career in context if we thought that his insights were both crucial to later developments and proposed originally for a different purpose. The assumption that his work was 'modern' but ahead of its time hinders our ability to understand its original context. But perhaps this is a task for the historian, not the biographer. With its wealth of detail, Orel's account certainly moves the study of Mendel's life onto a new level. □

Peter J. Bowler is in the Department of Social Anthropology, Queen's University of Belfast, Belfast BT7 1NN, UK.



Pendulums in practice

Paul Foulkes

My Own Right Time: An Exploration of Clockwork Design. By Philip Woodward. Oxford University Press: 1995. Pp. 176. £24.99.

To set out the general theory of the ideal pendulum is a fairly simple exercise in mathematics. The complications that arise when we examine how the device works in practice are altogether more intricate. Philip Woodward's book gives a splendid

of how to overcome the generally non-conservative damping forces we meet in practice.

To describe the quality of resonance in the resonator that constitutes the regulating mechanism, the symbol Q has been borrowed from electric circuit theory. This is the ratio of peak restoring force to peak damping force, a dimensionless number. The greater Q , the more efficient the resonator and therefore the more accurate the time-keeping. Woodward tells us that "for some reason horologists did not at first take kindly to it". The reason is surely clear: the inverse ratio, peak damping to peak restoring force, is intuitively more intelligible. The smaller that ratio, the better and more efficient the resonator. To achieve this, we must reduce the damping force, rather than indefinitely increase the restoring force. Hence $Q' = 1/Q$ is a quantity more easily visualized. To the mathematician this hardly matters, but it does to the practical mechanic.

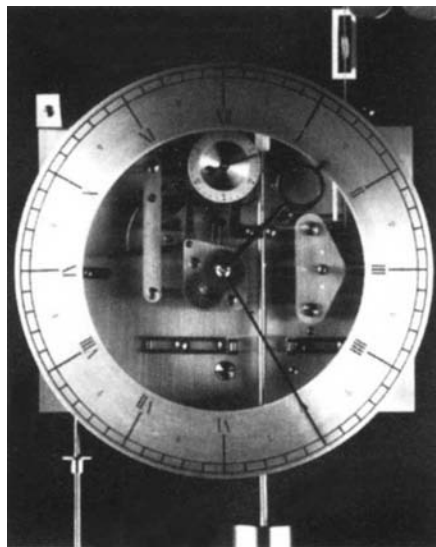
In trying to give an intuitive account of the pendulum's basic lack of isochronism, the author bids us to imagine a pendulum swinging half a turn and coming to rest upside down. This is not a happy sug-

gestion, because such a position of 'rest' is unstable; the pendulum would in fact topple over. The 18 per cent rise in period for a swing of 90 degrees is computable from the general solution of the pendulum equation, which involves an elliptic integral and is therefore rightly omitted.

The chapters on statistical error analysis and on flicker noise take the subject rather further than most ordinary accounts. Finally, there is some discussion of a source of error against which no mechanical remedy exists. This is the possible seasonal variation of the acceleration due to gravity at the Earth's surface, g . This can occur when the density in a layer of material below ground changes, for example when water is absorbed by porous rock in the rainy season.

Such changes are not precisely predictable: the user simply has to adjust his clock in the event. Changes in a pendulum's length, thermal disturbances and frictional losses in various mechanisms can be allowed for, but changes in natural 'constants' are much more elusive. □

Paul Foulkes is at 24 Granville Park, London SE13 7TEA, UK.



Clockwise — Woodward's remarkably accurate free pendulum clock W5.

account of this, with photos, drawings and graphs. While relating his own quest for an ever-more effective design for a mechanical time-keeper (his W5), he manages in clear and simple words to describe the main types of escapement mechanisms in clockwork right to the latest developments in mechanical art. For the purpose of keeping time, this has now been overtaken by the electronic revolution. Even so, much in mechanical horology remains both valuable and instructive in the field of applied mechanics.

The explanations of how these various devices work is supported by clear diagrams. To follow operations step by step is not always easy, but then the sequences are fairly complex, so that the reader must concentrate. The amount of mathematics is small and elementary, and calculus is avoided. This must have meant some self-restraint, since the author is an applied mathematician and physicist. Still, the effort is commendable, in view of the wide readership envisaged. Perhaps in a later edition the meagre appendix could carry some more theory. Meanwhile, we have here a study of great depth of the problems

Evolution of artificial intelligence

Terrence J. Sejnowski

Artificial Minds. By Stan Franklin. MIT Press: 1995. Pp. 449. \$30, £22.95.

WE are living in a postmodern age in which the functionalist architecture of artificial intelligence (AI), like the functional architecture of the Bauhaus, is beginning to look rather old fashioned. Stan Franklin takes the reader on a tour of the eclectic computational styles competing to replace the traditional symbol-processing paradigm that served as the unifying framework, and strait-jacket, for a generation of mind-modellers. He leads the mindful tourist on a whistle-stop tour of artificial neural networks, computational neuroethology, situated action, artificial life, chaos theory, quantum computing, neural Darwinism and a host of other hopeful approaches. Where symbolic AI was once the only game in town, there now seems to be one on every street corner.

A trend that has evolved over the past decade is a shift away from the toy worlds of early AI towards problems that arise from interaction with the real world. The first computers available to AI researchers had less computing power than a pocket calculator. So the problems attacked were highly simplified. Because computers are efficient at symbol processing, solutions were sought in the realm of logic. The world, however, is noisy and uncertain and does not come labelled by symbols, which

must be grounded in sensation and learned through experience. Inductive reasoning from incomplete knowledge is as important as deductive reasoning. Ironically, it is not logic but the theory of gaming that provides the mathematical framework for inductive reasoning in an uncertain world, and George Boole, in his 1854 book *An Investigation into the Laws of Thought*, which introduced Boolean logic, devoted the second half to the theory of probabilities.

Artificial Minds is organized around three debates in AI. The first is whether a computer can ever be programmed to think. This is a recurring question that does not seem to have a resolution, in part because the definition of thinking keeps changing. The problem has been revived recently by Roger Penrose, who doubts the possibility that algorithms could ever think like a mathematician.

I am highly sceptical of claims that something is impossible. A small change in an innocent-looking assumption can drastically change the conclusions. Some aspects of thinking have already been automated, but there are others, such as subjective feelings, that do not have an obvious mechanical analogue. There is a vast middle ground to explore and much of it can be approached with empirical studies of brains and the invention of ever-more clever devices.

The second AI debate, led by the con-

nectionist alternative to symbol processing, concerns Jerome Feldman's 100-step rule. The brain can solve many difficult cognitive problems in half a second. If a neuron takes 5 milliseconds to perform one operation, then it should be possible to devise algorithms for solving the same problems that require only 100 steps. Although the SOAR program of the late Allen Newell respected this constraint, most AI programs take millions of steps; connectionist algorithms, by contrast, distribute the information over many simple processing units and work in parallel. The first-generation neural network models, however, were weakest where symbol processing is strongest, in representing relationships between abstract categories. The analysis of nonlinear networks as dynamical systems is still in its infancy and I suspect that the next generation of network models, based on spiking neurons, may provide additional computational abilities that may overcome this limitation.

The third debate reaches into the heart of AI and questions the need for representations at all. Rodney Brooks incorporates a hierarchy of reactive behaviours into walking machines that inhabit the Massachusetts Institute of Technology; the emphasis is on action selection rather than highly detailed sensory representations. Of course, as the control of actions becomes more sophisticated, the burden of intelligence is simply shifted from the sensory to the motor side. Franklin declares his own bias for a theory of intelligence based on an action-selection scheme, in which tasks are solved by a variety of specialized agents. These autonomous agents remind me of programs on the World Wide Web, such as 'spider' agents that seek out and catalogue information available at Web sites. If the digital computer served as a model of the mind for a generation of students in cognitive science, will the next generation be guided by their experience surfing on the Internet?

Artificial Minds is a readable guide to the evolution of ideas in artificial intelligence. It is intended for a general audience that already has some familiarity with symbolic AI and could be used, like a tour guide, to decide what areas of *nouvelle AI* to visit in greater depth. My only disappointment is that cognitive neuroscience, the science of real minds, plays only a small part in the book, perhaps because this field is of less interest to the author, whose background is in mathematics and computer science. New techniques for probing brains in action are uncovering unexpected insights into the organization of human minds. But that is another tour. □

Terrence J. Sejnowski is at the Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.

Hard to swallow

Bruce Traill

The Food System: A Guide. By Geoff Tansey and Tony Worsley. *Earthscan*: 1995. Pp. 259. £15.95 (pbk).

PERHAPS it is unfair for an economist to review this book. Economists are mostly of the view that markets are a good thing and that the 'invisible hand' provides a cheap and efficient mechanism for allocating resources and rewarding effort. Of course, we recognize that markets are not perfect and that in certain cases political intervention is needed to regulate or otherwise remedy market failure: markets may become monopolized, 'externalities' may exist that are not taken into account by producers or consumers (such as the environmental implications of intensive farming or food packaging) and consumers may have inadequate knowledge to make informed choices (over food safety or nutritional issues, for example). The economist's response is to identify precisely the source of the market failure and take the necessary corrective action (provided that the benefits of doing so outweigh the costs).

The authors of this book start from the opposite position, believing that the market mechanism is fundamentally flawed and cannot be trusted to perform even the simplest tasks. They therefore conclude that the food system should be thoroughly managed to achieve a set of food-policy goals: the equitable provision of a safe, secure, sustainable, sufficient and nutritious diet for all. It is hard to argue with such basic objectives, the real question of course being whether politicians, civil servants and (much favoured by the authors) public-interest nongovernmental organizations (PINGOs, as opposed to BINGOs, business and industry NGOs) can do a better job than a lightly regulated market of telling businesses what food to produce, and where and how, and consumers what to eat.

The authors spend most of the book describing their interpretation of the way in which the food system works in the developed world (with examples primarily from the United Kingdom and Australia) and where it is going wrong. A short set of conclusions discusses "general approaches" to the development of food policies for a new millennium.

In the book's favour, it does provide a lot of facts on a very wide range of subjects from ecology ("food and the biosphere") and food-poisoning organisms, through industrial and retailing change, to advertising and consumer market segments and much more besides. The

book is divided into three main parts, the first dealing with global issues: it describes world agriculture (and its harmful effects on the environment); gives a short economic history of the world (and how it has resulted in a system in which rich-country citizens over-consume processed foods while poor-country citizens starve); and reviews nutrition science, food safety and the psychology and sociology of consumer food choice (and why we are all overweight and consequently exploited by clever advertising agents over our fears of not being slim and beautiful).

The second part of the book describes the principal actors in the food chain and their interaction. It covers farmers and the effects of agricultural policies on farming practices and farm income distribution, and the growing scale and global nature of the (often multinational) corporations involved in trading agricultural commodities, manufacturing food, sourcing produce and retailing and catering. It is not by this stage in the book a surprise to find that the authors believe that consumers have too little power in the system and that small-scale local operations are preferable to large-scale global ones.

Power is returned to as the main theme in the final part of the book dealing with what the authors term "food control" mechanisms. These are the use of science and technology, information and sophisticated management techniques by actors to gain control in the food chain, and the use of policy to regulate the power and determine the distribution of benefits among the actors. There is the familiar argument that large corporations influence the allocation of research funds and so control the direction of technical progress. They also pay for advertising foods (and drinks) that are nutritionally harmful, and fund food-related (mis)education in schools. To top it all off, policies that could moderate these effects are formulated today by civil servants with male-centred ideologies and, according to an Australian survey quoted, a "pseudo-religious belief in the market".

Derek Cooper, a presenter of the British radio series "The Food Programme", is quoted on the front cover as saying: "if you want to understand why the world feeds itself the way it does, you must read this book". Be warned that, although one can have sympathy with many of the points the authors make, and despite an impression that the authors are often trying to appear impartial, the understanding one is fed is not entirely balanced. □

Bruce Traill is in the Food Economics Group, University of Reading, Reading RG6 6AR, UK.

The Moon's a balloon

Henry Gee

Apollo 13. A film directed by Ron Howard. *Universal Pictures: 1995.*

Lost Moon: The Perilous Voyage of Apollo 13. By Jim Lovell and Jeffrey Kluger. Houghton Mifflin: 1994. Pp. 378. \$22.95. Published in paperback as *Apollo 13*. Pocket Books, \$6.50; Coronet, £5.99.

The Apollo Adventure: The Making of the Apollo Space Program and the Movie Apollo 13. By Jeffrey Kluger. Pocket Books/Boxtree: 1995. Pp. 199. £8.99 (pbk).

THE most remarkable thing about the Apollo 13 mission is that it got anywhere near the Moon at all. The veteran astronaut Jim Lovell (played by the deservedly ubiquitous Tom Hanks), commander of the ill-starred spacecraft, works out a course correction, by hand, with pencil and paper. Not quite certain of the correct answer, he asks mission control to verify his calculations. More than 200,000 miles away, men in short-sleeved nylon shirts and unfeasibly thick spectacles gather round a console and one of them goes over Lovell's maths — with a slide rule.

The story of Apollo 13 is simply told. On Monday 13 April 1970, while proceeding in a moonwardly direction, an electrical fault sparks an explosion in a tank of liquid oxygen, rupturing the spacecraft's insides. Deprived of much of the oxygen needed for their fuel cells and (let us not forget) breathing, the three astronauts shut down the command module and take refuge in the lunar module, a short-range bicycle built for just two (and for those, cosily). Any thoughts of landing on the Moon are cancelled as crew and controllers alike struggle to bring the spacecraft home safely. Which they do, of course.

Not that this foregone conclusion should deter one from seeing *Apollo 13*, a thoroughly involving story, far from the *Boy's Own* adventure that one would almost be bound to expect. The universally fine acting, the sharp, documentary-style direction and (one has to say) attention to period detail hold one's interest throughout. When the astronauts wipe condensation off the windows so they can see the Moon, you are there with them, willing them to come home safely.

But *period* detail? Surely: this is 1970, the year the Beatles split up (a cause of much *angst* for one of Lovell's teenage daughters in the film), when even quite normal people wore flared trousers, plastered brown and orange hessian up the walls, and lava lamps were actually in. The film opens with the landing of Apollo 11, as seen on television in the Lovells' lounge, the décor of which is

simply hideous. In 1970, I was eight years old, and British coinage was yet to be decimalized (I remember saving 5s 11d to buy a plastic model kit of the lunar module from Woolworth's). Digital



NASA's soft under-belly — Apollo 13 astronauts during lift-off in a scene from the film. Left to right: Bill Paxton (playing Haise), Kevin Bacon (Swigert) and Tom Hanks (Lovell).

watches hadn't been invented, Clive Sinclair had still to popularize the electronic calculator, and NASA scientists boasted of a computer so powerful that it could store a million bits of information yet be fitted into any convenient ballroom.

Given such primitive technology, what business did we have even to tip our hats at the Moon? At the time, NASA folk talked publicly of science, exploration and pioneering spirit. When asked by a congressman why, when Apollo 11 had already landed on the Moon, one would need to go back, Hanks (as Lovell) responds that Christopher Columbus had had no such qualms. People talked cheerfully of regular trips to the Moon and the establishment of a permanent lunar base: and yet, so far, barely a *minyan* has trod the lunar surface, and the Moon has had no return bookings since Apollo 17 in December 1972, more than 20 years ago.

Not that anyone, apart from small children, was under the illusion that the manned space programme was anything other than a dig at the Soviet Union. Of the dozen moonwalkers, all have been

Americans. In this political respect, Apollo was a complete success. In that we have not returned, and the sum total of scientific knowledge gained by the Apollo programme was not significantly greater than nil, it was an abject failure. NASA's greatest scientific successes have come, as they always have, from unmanned probes. As Bruce Murray pointed out in his book on the Voyager missions, it is a sad irony that Voyager 2 reached Uranus on the same day that the space-shuttle Challenger exploded after barely clearing the ground. Now, when even the shuttle-delivered Hubble Space Telescope is occasionally upstaged by more advanced telescopes on the ground, the continuing political devotion to the shuttle programme and the space station is a mark of an agency whose starship doesn't quite make warp speed.

This, then, is the real tragedy of the manned space programme, and why the film is so affecting: that so many people, on the ground as much as in the spacecraft themselves, should expend so much loyalty, effort and commitment, and put their lives at risk, to an end that was never more than tawdry and is now completely pointless. That they managed to get as far as the Moon inside what, to modern eyes, looks like an over-engineered tumble-dryer only embellishes their quixotic bravery.

To take the longer view of the confirmed science-fiction reader, the Apollo missions were simply ahead of their time, like Victorian designs for steam-driven aeroplanes — the miracle being that with Apollo the designs often worked. Perhaps, one day, when we are very much richer and have worked out a genuine rationale for the enterprise, something like the space station will become a viable platform for manned exploration of the Solar System, perhaps even the stars... one day.

Back on Earth, *Lost Moon* is the book that inspired the film, Lovell's personal account of his part in the space programme, with editorial help from the writer Jeffrey Kluger. Lovell had flown twice in the Gemini missions (the precursor to Apollo) and had orbited the Moon in Apollo 8. The Apollo 13 mission was to have been his swansong, when he would finally set foot on the Moon.

It was not to be, but his personal account of how he made his way into the space programme by way of the US Navy (in which he had been a test pilot) is interesting reading. The book is written like a novel, in the third person, and the better parts remind one of James

Michener's weighty epic *Space* (particularly the scenes at the US Navy test centre at Patuxent River, Maryland, which presumably drew on a common fund of experience). Matters sag, somewhat, as Kluger displays the extent of his researches with interminable reconstructed mission-controllathons. Here is one:

"Ten minutes to burn," Haise announced. Shortly afterward he called, "Eight minutes to burn," then "Six minutes to burn," then "Four minutes to burn." Finally Brand, at his Capcom station, took up the call.

"Jim, you are go for the burn, go for the burn."

"Roger, I understand," Lovell said. "We are go for the burn."

Somehow the film manages to convey the technical niceties of the mission without such tedious exegesis. Nevertheless, *Lost Moon* is packed with information on a mission that NASA would probably

rather we forgot about, and jackdaw space enthusiasts will not be deterred by the occasional attempts at literature.

The Apollo Adventure (the book of the film of the book) eschews even that. It is little more than a scrapbook of mission details, accounts of the filming and assorted Apollo miscellany that Lovell and Kluger were unable to cram into the earlier book. These include transcripts of astronauts bitching about their employers, no more interesting than one would expect for them having been classified until recently. The design and production are tacky: the random pepperings of spacecraft schematics would be great were one the kind of person (God forbid) who absolutely has to know the location of every storage compartment in the command module—except that the printing is too indistinct to read. Save your money and see the film instead. □

Henry Gee is an assistant editor of *Nature*.

better than "constancy" and "balance". They will not delight in their characterization as the smart heroes who oppose the ignorant, evil managers of national parks. With hindsight, ecologists might have decided to do things differently at Yellowstone or Tsavo. Nevertheless, those who manage fire or large mammals in parks that are token remnants of the original native ecosystems have awesome tasks to perform.

Finally, we meet the theoreticians, still waiting for a modern Newton to give them a grand synthesis. Budiansky writes at length and fairly accurately about grouse, fisheries, the restoration of oak-savannas and which species one cannot remove from food webs without dire consequences. Those expecting victims will not be disappointed. Edward O. Wilson is described as "a relentless popularizer" whose conclusions "are something closer to politics than science" and as the known collaborator of "misanthropic" Paul Ehrlich, the "over-population guru".

The subject is extinction rates. We are told to be surprised that the observed rate "is not 50,000 species per year, but *one* species a year". Only those who read footnotes will find that Budiansky derives this estimate from a sample of one in 2,000 species. So the yearly total is in fact 2,000 extinctions—a rate that will surely increase at least tenfold in the coming decades for reasons Budiansky discusses. A yearly rate of 20,000 extinctions is within a factor of three of Wilson's much-derided estimate, yet four orders of magnitude greater than the rate implied in the main text.

The arithmetic also fails when Budiansky discusses the relationship between the number of species an area contains and its size. This is the key concept underpinning

predictions of the extinctions that follow habitat losses. He quotes an "unintended parody of the entire concept" in a novel whose hero finds 200 beetles within a square mile, then estimates that within a day's drive there will be 500,000 species. Knowing that species numbers increase under such circumstances in proportion to no more than the one-tenth to one-fifth power of the area, ecologists predict 650 to 2,100 species within a 200-mile radius. Far from being a parody, their predictions (unlike the hero's) are reasonable. So too are their estimates of species losses following habitat destruction.

For someone who can write engagingly and some-

Nature lovers and other villains

Stuart L. Pimm

Nature's Keepers: The New Science of Nature Management. By Stephen Budiansky. *Free Press/Weidenfeld and Nicolson*: 1995. Pp. 310. \$25, £18.99.

ONE judges a book not by its cover but by the quotations that adorn it. Ennobled by the statement on the back of *Nature's Keepers* that "modern ecological research is providing the tools for effective nature management", ecologists ought to ask for more money. Yet the volume also carries praise from Gregg Easterbrook, author of *A Moment on the Earth*, a deeply flawed book that spawned a vigorous Internet-based competition this summer to see who could find the most scientific errors per page. So is *Nature's Keepers* another book trashing science, conveniently timed to appear just as the US Congress is contemplating reversing its long and impressive record of protecting nature? The text is as divided as its cover.

In "the cult of the wild", the book impales nature lovers who revere William Wordsworth, the poet, or Chief Seattle, the American Indian ecological prophet. Wordsworth did not want his daffodils appreciated by the untutored, working-class hordes that a proposed railroad would bring. And Chief Seattle's famous speech—"Earth does not belong to man, man belongs to Earth"—is a fabrication of a modern film script. He evidently said nothing about the environment. Debunking myths is wicked fun. It also misses the point. Christmas is still deeply significant to Christians who know that decorated trees are a Victorian

import, that 25 December is suspiciously close to the date of pagan winter-solstice rituals and that astronomical explanations for the Star of Bethlehem (comets, supernovae) came in the wrong month of the wrong year. Historically, Amerindians may have exterminated most of the large vertebrates they encountered, but we should not denigrate their modern religious beliefs about the environment.

Next comes "the science (but mostly politics) of nature management". Ecologists are praised for knowing that "disturbance" and "change" typify nature



From *The Grizzly Bears of Yellowstone: Their Ecology in the Yellowstone Ecosystem, 1959–1992* by J. J. Craighead, J. S. Sumner and J. A. Mitchell. Island, \$100.

times incisively about environmental issues, Budiansky consistently misses the big questions. There are difficult ethical and economic choices in how we value biological diversity. Those who manage this diversity face not only scientific uncertainty but also the knowledge that their actions may take decades to unfold, time in which society's values may change. Scientists who predict high future extinction rates are not omniscient, but believe, simply, that ignorance is not bliss. How sad that Budiansky misses the chance to write thoughtfully about the interfaces between nature, ethics, economics and our future. □

Stuart L. Pimm is in the Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996, USA.

Ways of seeing

David Knight

A History of Scientific Thought: Elements of a History of Science. Edited by Michel Serres. Blackwell: 1995. Pp. 760. £75, \$100.

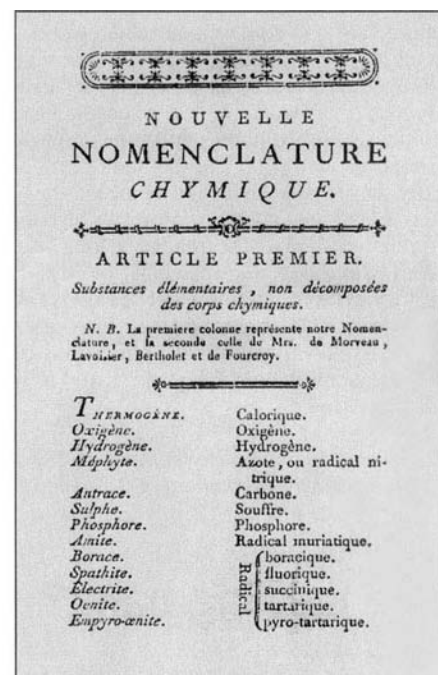
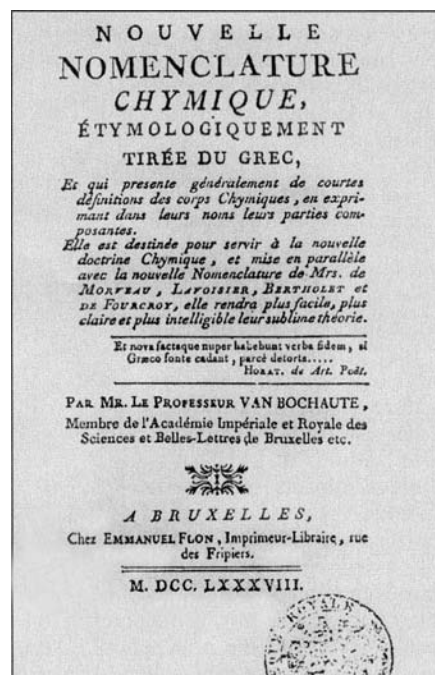
THEY order things differently in France, and it is good for outsiders to see how things look there. Although the dust jacket has a portrait of Michael Faraday in his laboratory, inside he makes only a passing appearance in one sentence (like God in the *Origin of Species*); the book does not cover everything. It is striking that we do not get to Galileo until page 280, and to the nineteenth century until about page 400: despite the exponential growth of science, the focus is on beginnings. The book is not intended as a work of reference; rather, it is a series of meditative or considered essays, examining nodal points in the long history of science from the first emergence of experts writing on clay in Babylonia.

Michel Serres himself evokes the beginnings of geometry in Greece and then in an impressionistic *tour de force* calls up the Paris of the years around 1800 when it was the centre of excellence in all the sciences. He sketches the emergence of scientists as a kind of secular clergy; the shade of Auguste Comte indeed stalks through the book, and we learn about his calendar where the saints are men of science, as well as the Revolutionary calendar with its ten-day weeks. Serres is bold enough to paint for us the dazzling big picture, the general plan, whereas most Anglo-Saxons, perhaps in our Baconian tradition, are more like bricklayers than architects in constructing the historical edifice.

Where it comes off, this is very splendid; and the various authors do their best to show how the events they describe are

nodal, rather like the poet-scientist J. W. Goethe's Ur-phenomena. They do their best also to show the unexpected, the non-teleological in the history of the sciences. We know that if Humphry Davy had not isolated potassium when he did, J. J. Berzelius would have done so within a year or two; and if he had not subsequently declared the elementary nature of chlorine, then J. L. Gay-Lussac would soon have done that. There is something like inevitability in the development of science;

1789–1914, where specialization and exponential growth were the order of the day, the focus is on individuals: A. L. Lavoisier, Charles Lyell, Gregor Mendel and D. I. Mendeleev. These are not surprising names, except perhaps that they include no classical physicists; there is no attempt here to revise the pantheon. Geof Bowker gives us a close textual analysis of Lyell's writing; Bernadette Bensaude-Vincent brings her usual clarity and sharpness to her studies of the chemists, placing them carefully in con-



Cover and page of the Flemish chemist Karel van Bochaute's *Nouvelle Nomenclature* (1788). From *Lavoisier in European Context: Negotiating a New Language for Chemistry* edited by Bernadette Bensaude-Vincent and Ferdinando Abbri. Science History Publications, \$45.95.

but there is also the contingent, as these authors emphasize. Close examination of social context and of personal careers certainly diminishes our belief in progress and inevitability: we all know how important opportunism and career development can be. These authors take history seriously.

We find here some excellent storytelling; for example from Michel Aulthier on Archimedes, warfare and science, and from Bruno Latour on Louis Pasteur, Felix Pouchet and spontaneous generation. Others give us lucid straightforward accounts without many surprises, such as James Ritter on early mathematics, Paul Benoit and Françoise Micheau on Arab science, or Jean-Marc Drouin on naturalists and travellers. Some wear their learning lightly, such as Catherine Goldstein writing elegantly about mathematics moving from marginalia by amateurs to journal articles by professionals as *Homo ludens* gives way to *Homo faber*. Some, such as Isabelle Stengers, have written something like essay reviews for a learned journal, in her case about recent studies of Galileo, which are closer to meta-history than to narrative.

For the long nineteenth century,

text — as does Drouin with Mendel, where reality and myth are particularly intertwined. These are 'exemplary' essays giving us some clues in our approach to their century; but in contrast to its close studies of the early and mediaeval periods, the book is thin in its coverage of the time when disciplines and professions were emerging, and the frontiers between sciences were being established and policed. Societies, journals, university departments, medical schools as midwives of biology and chemistry, great exhibitions, lectures stimulating the public understanding of science, addressed to the élite or to working men — we learn very little of all this.

And for the twentieth century, we have a close study of devices for the oil industry; an interesting story of Frédéric Joliot and the intersection of science with politics and war; and a carefully contextual history of the computer by Pierre Lévy emphasizing the different roles played by different people with different ends, in a complex story in which one should not look for some crucial and datable foundation-moment. Here, as in the previous chapters, this approach keeps the book open for the

reader without a modern scientific training; as an account of the science of our century it clearly does not give any kind of coverage, but it does indicate seminal points. Joliot on atomic weapons has contemporary resonance, but it is perhaps surprising that there is no sequel to the naturalists or to Mendel, given the importance for us of ecology and genetics.

We thus have a book full of stimulating insights, something to be recommended to students or to the general reader who has some idea of what science is and was about. We find some memorable episodes such as the mediaeval attempt to make theology a science, as described by Paul Benoit; but probably for scientists, the emphasis on beginnings and times remote from our own will make the book particularly desirable. These lively accounts — and the anonymous translator is to be congratulated — take the reader a long way from the brief run-throughs of recent history to be found in science textbooks or encyclopaedias. And the book also gives professionals in history and philosophy of science something to engage with; it is very welcome. □

David Knight is in the Department of Philosophy, University of Durham, 50 Old Elvet, Durham DH1 3HN, UK.

The last big bang?

Mike A'Hearn

Rogue Asteroids and Doomsday Comets: The Search for the Million Megaton Menace that Threatens Life on Earth. By Duncan Steel. Wiley: 1995. Pp. 308. \$24.95, £16.99.

WILL we or our descendants all die when an extraterrestrial object comes hurtling down onto Earth? This question was brought to the fore by the enormous publicity surrounding the collision of the fragments of comet Shoemaker-Levy 9 with Jupiter in July 1994. Even before these events, the question had stimulated some astronomers to propose searches to identify all large asteroids that might hit Earth in the foreseeable future. The question had also stimulated developers of nuclear weapons to propose mechanisms for deflecting such asteroids before they collide with Earth. At the behest of the US Congress, NASA has appointed several committees to consider detection and deflection of potentially hazardous asteroids, the most recent of which, the Shoemaker Committee, has just released its report. Meanwhile, the International Astronomical Union has also appointed a working group to consider what actions are scientifically appropriate in studying and detecting

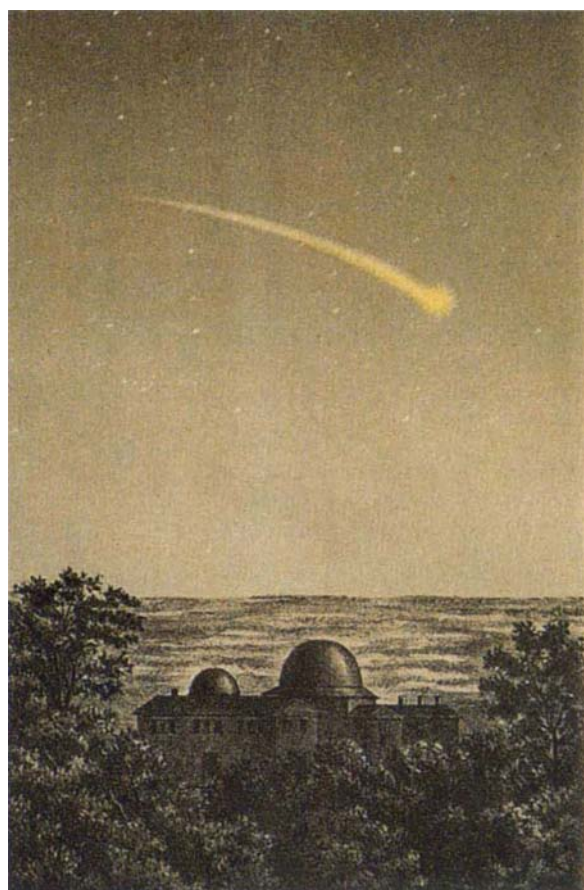
near-Earth objects.

As someone engaged on the periphery of this issue, it seems clear to me that certain aspects of the problem have been investigated thoroughly and are well understood, whereas others have been almost ignored and so are not understood. It is technically possible to carry out a search, on a timescale of ten years or so, to detect 90 per cent of the near-Earth asteroids with radii greater than 1 km that might hit Earth in the foreseeable future.

On the other hand, the hazard due to long-period comets is poorly understood and we do not know how to detect these more than a year or two in advance. Similarly, we do not fully understand the nature of the hazard even from asteroids of a given size. Studies of last year's Jupiter impacts (discussed in a lengthy epilogue to the book) show the importance of a process previously not even considered: when material ejected during the impact falls back down onto the atmosphere, it dramatically heats the atmosphere over very large areas.

It is also unclear how the public views the risk of an extraterrestrial impact compared with that of, say, an aeroplane crash. Statistically, the risks of death from the two causes would seem to be of the same order of magnitude. Extraterrestrial impacts are, of course, much less frequent but have a dramatically greater effect, leading to a huge uncertainty in the actual number of deaths as well as a notable difference in public perception of the risk.

Steel aims to heighten public awareness of the risk and thus develop political support for funding the searches and other studies that might ultimately allow us to save ourselves or our descendants from a worldwide catastrophe. He writes about facts that are accepted by most astronomers and ideas that are wildly speculative, as well as about everything in between. There are vast uncertainties in many of the predictions and he points out many areas where astronomers are in disagreement. Examples of past impacts include the generally understood small impacts, such as Meteor Crater in Arizona and the explosion above Tunguska in 1908, but Steel also writes about other impacts, the effects of which are not so widely accepted. It seems to me that most



Mary Evans

Passing threat — "The Meteor", by Guillemain (1877).

astronomers accept the theory that the impact that created the Chicxulub crater in the Yucatan also caused the extinction of the dinosaurs at the end of the Cretaceous period, although this is apparently less widely accepted among palaeontologists, some of whom still argue for a more gradual extinction. Unclear to both astronomers and palaeontologists is whether large craters and extinctions are really periodic.

With its wide range of speculation, every scientist, whether astronomer or geologist, palaeontologist or archaeologist, will disagree with one or another of the suggestions in this engaging and worthwhile book. Steel's proposal that the Taurid meteor showers sparked off the building of the first stage of Stonehenge will certainly charge up many readers, some positively and others negatively. This is, of course, his purpose in writing the book — to encourage others to think about the problem. The basic question is whether we should be terrified at the prospect of a large impact. Should we spend the money now to begin the search for potential impact objects and find out more about this particular hazard — a hazard that is undoubtedly real but of uncertain magnitude? □

Mike A'Hearn is at the Astronomy Program, University of Maryland, Stadium Drive, College Park, Maryland 20742, USA.

Once a cigar, always a cigar

Peter J. Swales

Why Freud Was Wrong: Sin, Science and Psychoanalysis. By Richard Webster. HarperCollins/BasicBooks: 1995. Pp. 673. £24.99, \$35.

ALTHOUGH touted on the jacket as a comprehensive biography, this is not a book that enriches or even attempts to digest our factual knowledge of Freud's life and work. Rather, availing himself of a variety of critical viewpoints advanced in recent years by other authors, Richard Webster seizes on several telling episodes in the origins and history of psychoanalysis to engage in a relentless polemic that, if at times flawed in its simplifications, is nevertheless lethal in its total impact.

At the heart of the author's revisionist synthesis lies a provocative though tenable appraisal of Freud the person. Several writers have previously pointed out a messianic element in Freud's character, but Webster sees it as the very animating principle that goaded Freud onward. Inured from earliest childhood to fulfil his parents' grandiose expectations, Freud sought — and of course found — the kind of secular immortality to be attained only by instating oneself as a world-redeemer.

Consumed by such a vanity, Freud was a scientific somnambulist, blinded to empirical reality by his own wishful agenda, oblivious to what did not match his own preconceptions and ready to appropriate anything that would help to support his madcap speculations — a self-deceiver incapable of recognizing the fallacies he perpetrated and the follies he propagated.

According to Webster, Freud's descent into folly began during his early years as a physician when, credulously submitting to the misguided notions of his mentors, Jean-Martin Charcot and Josef Breuer, he persistently misdiagnosed as psychogenic an array of mysterious somatic symptoms that neurologists today would not fail to recognize as indicative of underlying organic pathology — closed head injuries, temporal-lobe epilepsy, syphilis and so on. Positing a discrete neurophysiological syndrome of a psychological origin, often following traumas, Freud routinely deemed such symptoms to be hysterical — by Webster's reckoning, a grab-bag diagnosis of no objective merit that served merely to mask the then profound neurological ignorance.

Constrained by the limits of the knowledge of his epoch, and without the benefit of modern medical equipment, Freud surely committed such blunders often

enough. But the author fails to give due consideration to alternative explanations, such as the real possibility that many of Freud's early patients were simply casualties of life, whose somatic symptoms he then managed to vanquish with words only because they were products of make-believe in the first place.

Take, for example, the little-observed



Illustration by Ronald Blommestein

“talking cure” was actually responsible for their disappearance.

By gainsaying Freud in such a take-it-or-leave-it fashion, and by ruthlessly exposing many empirical and logical absurdities in his claims, Webster generally makes a good case that Freud went awfully wrong. He is far less successful when he seeks to belittle Freud by minimizing his originality, for here the reader is offered only wearisome variations on a theme. According to Webster, the tenets of Freudian theory are merely new editions of select facets of the Judaeo-Christian tradition — so psychoanalysis is ultimately explained away as being no more than a disguised and secular version of Western theology, with Freud, of course, as godhead.

From childhood on, Freud was indeed preoccupied with biblical motifs, and their influence on him is not to be underestimated. Before long, however, Webster's attempt to frame Freud's thought-system in this context begins to ring hollow for, in truth, his interpretation is merely a substitute for a closer appreciation of Freud's wide-ranging sources.

If mankind is truly to be cured of psychoanalysis — and if we are all to be cured of having repeatedly to contradict Freud — then we need to understand adequately the man and his creation. Some familiarity with the history of ideas in nineteenth-century Germany and Austria is therefore indispensable. But Webster's frame of reference, as reflected by his literary quotations and bibliography, is so Anglo-centric as to provoke in the more attuned reader an acute sense of cultural dislocation.

Webster portentously disposes of Freud's most intellectually formative years, his decade of studies in biology and medicine, in half a page. He also eschews dozens of Freud's published letters from that period that identify thinkers of seminal importance in the shaping of his *Weltanschauung* — Gotthold Ephraim Lessing, Johann Wolfgang von Goethe, Jean Paul, Ludwig Feuerbach and Ernst Haeckel, to name but a few. Inevitably, such remissness makes the subsequent inception and evolution of psychoanalytic theory more difficult to comprehend.

Freud would claim that his psychosexual theories were compelled by clinical observations — that, reluctantly, he had been forced to confront sexuality in investigating psychopathology on his couch. Following Frank Sulloway, Webster rightly emphasizes how Freud's theory-building after 1896 was mainly inspired by assumptions smuggled in from contemporary biology.

fact that, on termination of her treatment, Breuer's patient Anna O. — the subject of the legendary ‘first case’ of psychoanalysis — confessed that she had simulated her symptoms. But Breuer, who had devoted himself to her care daily for most of 18 months, would hear none of it; and he would later construe her admission as consistent with a cure of her hysteria. Webster, tenacious in his organic absolutism, disregards this ominous episode.

Although not himself a man of medicine — his own background is in English literature — the author does not doubt that Anna O. was afflicted by a severe neurological disorder, yet he supposes its array of florid symptoms underwent spontaneous remission one after another. Moreover, because the patient had naturally discussed all of these with her physician, she and Breuer — and, later, Freud too — fell into the *post hoc* fallacy of supposing that the

In particular, Webster stresses the importance for Freud of Haeckel's "biogenetic law" — the by now long-discredited principle that ontogeny recapitulates phylogeny. But, oblivious to Freud's background in biology — it was in fact Haeckel's brand of Darwinism that had stimulated Freud to take up studies in science in the first place — Webster supposes it was Wilhelm Fliess, Freud's biologist friend, who equipped him, late in the day, with such a viewpoint. And this helps to obscure a much larger truth.

Webster is astute in perceiving as fundamental to Freudian interpretation what he dubs Freud's "theory of correspondences" — a mode of ontogenetic reductionism accomplished by appeals to homologies and analogies whereby, in Freud's scheme of things, every silk purse can inevitably be unmasked as nothing but its "prototype" or "imago" — that is, a sow's ear. And he recognizes a basic epistemological kinship between the heuristic styles of Freud and Fliess — for, throughout his writings on biorhythms, Fliess also uses an audacious reductive methodology based on the principle of the prototype (Webster's "lowest common denominator").

But the author supposes that Freud arrived at this mode of interpretation in his clinical practice in the years after 1910, whereas in fact it had by then long been the quintessence of the Freudian style; and he fails to appreciate the origins of this approach in the unconstrained morphological conjecture characteristic of comparative anatomists and zoologists in Germany from Goethe to Haeckel — luminaries whose works were fundamental in shaping the thought systems of Freud, Fliess and several of their late nineteenth-century scientific contemporaries.

Historically speaking, the doctrines of Freud and Fliess are properly seen as bastard children of nineteenth-century morphological science. But this means, in effect, that both men's thought systems are throwbacks to the *Naturphilosophie* of the Romantics and the "natural supernaturalism" of the German idealist philosophers (re-tailored for the Anglophone world by Thomas Carlyle, the Alpine Christian). And it is no coincidence that basic to their programmes was precisely the sort of secularization of biblical motifs that Webster here attributes to Freud.

A less polemical and more informative "critical intellectual biography", then, would trace the materialistic pantheism of Freudian theory to its more immediate sources (Carlyle among them). And such a book — one that would inevitably comprehend Freud's life as one actually lived on the model of a *Bildungsroman* — might well vindicate Webster in his identification of Freud as Messiah, showing him to be more on target in some of his intuitions than even he himself knows.

With its trenchant assault on the empirical and logical foundations of Freud's theory-building and its cogent depiction of psychoanalysis as a religion, Webster's book, albeit largely derivative, stands as a valid and important addition to the literature. Its value as a contribution to history is vitiated, however, by the author's lamentable ignorance of the cultural and scientific contexts in which Freud worked and of which he was very much a product, and his wanton disregard of much of the primary material that illuminates his intellectual genealogy.

Emboldened by his demolition work and concerned to write more than just another polemic against Freud, Webster ends by trumpeting an oddly messianic note of his own — that out of the ashes of psychoanalysis a new psychology can now arise on the wings of neo-Darwinism. But his phoenix fails to take flight, for he omits to say what such a psychology might consist of and to what epistemology it might submit. Furthermore, in the wake of the demise of both Marxism and Freudianism, to invest in the basic principle that, with time, through natural selection, a sow's ear can become a silk purse is surely to bet against the odds — this at the very least. □

Peter J. Swales, an historian of psychoanalysis, is at 285 Mott Street, New York, New York 10012, USA.

Ancestral patterns

Jane Maienschein

Biology Takes Form: Animal Morphology and the German Universities, 1800–1900. By Lynn K. Nyhart. University of Chicago Press: 1995 Pp. 414. \$75, £59.95 (hbk); \$27.50, £21.95 (pbk).

TWENTY years ago, Garland Allen published his *Life Science in the Twentieth Century*. The first chapter included a seven-page section on "The Science of Morphology", which Allen defined in terms of the study of living forms, tracing common evolutionary ancestors and using the comparative method to reconstruct phylogenies. Ernst Haeckel and August Weismann led the way in a field characterized by speculative theorizing. Then, according to Allen, a younger generation rejected this speculation and phylogenizing, producing a "revolt from morphology". In turn, a younger generation of historians challenged Allen's avowedly Marxist insistence on revolution and suggested a more gradual, evolutionary interpretation of change.

While this discussion was developing, Lynn Nyhart was an undergraduate at Princeton, going on to graduate school at the University of Pennsylvania. She felt

that neither Allen's short sketch nor the responses were capturing the great diversity of German morphology. Never an established university discipline, morphological study nonetheless gained a different kind of prominence in German intellectual discussions than either Allen or his critics (who focused on American biology and its German roots) had fully appreciated. Rather than attacking these interpretations, Nyhart instead convincingly shows that we need to go further to explore more of what had been the lively study of morphology — and that it matters for our understanding of science. By offering her compelling study in the spirit of extending scholarship rather than demanding that her own interpretation replace all others, she reveals a healthy open-mindedness. The overall result should be an improved understanding of the scientific enterprise and how to study it. Her book is to be applauded for its good sense and provocative suggestions for valuable new research directions, and should stand as a worthy companion in the German study of form to classic explorations of Darwinian evolution, medical anatomy and embryology.

As Nyhart points out, recognizing the contributions of dozens of researchers as they moved among German universities over a century yields a different perspective than the study of only a single person, place or set of scientific contributions at a particular time. Although she thus lacks detail about the scientific contributions, she sees generational shifts that would otherwise remain invisible. She recognizes that older generations of scientists do not die as quickly as generations of new researchers and approaches appear, and that generations thus overlap intellectually rather than later generations neatly replacing earlier ones. The replacement may be further confounded by lack of institutional opportunities.

In the nineteenth century, E. S. Russell's three stages of morphology prevail (as elaborated in his 1916 book *Form and Function*): idealism gives way to evolutionary morphology, then to causal experimental analysis. Yet the stages overlap, and the intellectual successes do not map neatly to institutional successes. For example, Haeckel gained prominence for his public espousal of Darwinian evolution or materialistic monism, yet never established an effective research 'school' to produce students to follow and spread his 'truth'.

Nyhart touches on several intriguing questions. How can we explain the intellectual success of the morphological programme when it was never refined as a discipline with university chairs or journals in Germany? Why did Haeckel and his close friend Carl Gegenbaur look like the leaders when they had few students and little direct institutional influence in

Germany? And what does this case tell us about the nature of science and how we should best study its development?

Nyhart provides some answers. Morphology found niches in other successful disciplines, notably anatomy and scientific zoology. Haeckel stood out partly because of his popular approach and his assertive, indeed bombastic, style, and Gegenbaur developed an exceptionally loyal family-like following among his students, many of whom he placed in foreign universities where they could carry on his research approach and spread his reputation. Finally, we learn that science is done by individuals and embedded in a variety of institutions, in ways that deserve more careful attention, as do the timing of innovations and the generational and power relationships.

We have heard such messages before, but Nyhart is no simple constructionist who thinks that institutions and their context directly *cause* the science. Rather, the texture of science is rich and varied. We

need to explore more aspects, and implicitly all accessible aspects. Along the way, we will even learn much that is instructive for science today, such as the changing relationship of teaching and research, the emphasis on innovation and the competition among epistemological and methodological approaches.

As one of those who challenged Allen's interpretations by calling for a closer study of a wider range of research, I find Nyhart's study exciting, even though she points out that my own emphasis was itself limited in focus and even though she could benefit from a comparison with American and British studies of morphology. She is right: we need a rich diversity of studies in the history of biology, including explorations of odd niches from all sorts of perspectives — just as she has persuasively revealed that the science itself is rich and varied. □

Jane Maeinschein is in the Departments of Philosophy and Zoology, Arizona State University, Tempe, Arizona 85287-2004, USA.

White heat of medical practice

W. F. Bynum

Technology in the Hospital: Transforming Patient Care in the Early Twentieth Century. By Joel D. Howell. *Johns Hopkins University Press: 1995. Pp. 341. \$47.50, £39.95.*

ONE modern answer to the old question "Is medicine an art or a science?" might be "Neither, it is a technology". The medicine of the late twentieth century is certainly identified with, and surrounded by, machines, and many of its most dramatic aspects, such as open-heart surgery, nuclear magnetic resonance imaging, haemodialysis or incubators for premature babies, are more the products of technological realizations than of conceptual transformations. Both the power and the cost of medical care have been irrevocably touched by technology. Several of the innovations supposed to reduce costs, such as keyhole surgery with short hospital stays, also depend on recent and promised future technology.

If technology is simultaneously one of medicine's problems and one of its solutions, it is unlikely to disappear from the medical scene. It is therefore surprising that historians of technology and historians of medicine have generally had so little to say to each other. So Joel Howell's monograph is to be welcomed as a study that takes as its fundamental theme the birth of the modern technologically orientated hospital. Take away hospitals, and most of the really expensive equipment of contemporary medical care would disappear. Why and how did these institutions

respond to the lure of the machine? Did it make any difference to their patients?

Howell's study is rooted in the patient records of two particular institutions, the Pennsylvania Hospital in Philadelphia and the New York Hospital in New York City, between 1900 and 1925. His analysis of a statistically significant sample of these records allows him to ask several questions about the nature of hospital medical practice in these decades, which, he convincingly argues, coincide with fundamental transformations. The patient records bear eloquent testimony to change. They more than double in median length, and, more importantly, increase by more than five times in the average daily space a record occupies. This discrepancy highlights the reduced average hospital stay, a result heavily influenced by the coming of routine tonsillectomies. In 1900, 2 per cent of patients admitted to the Pennsylvania Hospital were suffering from diseases of the tonsils; in 1925, more than a quarter were. Similar figures obtain for the New York Hospital. Tonsillitis became the most common hospital disease, and the operation by far the most frequent surgical procedure. It first brought children into general hospitals as an ordinary occurrence, almost as a childhood rite of passage.

Howell's quantification of the ton-

sillitis phenomenon will probably come as no surprise to many readers with early memories of going under from ether or of ice-cream after surgery. It is almost a byway for Howell, because these children often generated no laboratory tests, such as a urinalysis (almost routine protocol by the end of his period), blood counts (common but still used selectively), bacteriological studies or X-ray examinations. The records of other hospital patients, however, contain the results of these and similar investigations with increasing frequency; often, too, data such as body temperature or pulse rate were collected serially and represented graphically. Blood counts became important for the diagnosis of typhoid fever or the prognosis of pneumonia, and following the discovery of insulin in 1922, the measurement of urinary sugar took on new significance. Printed forms helped to standardize the collection and reporting of data, and the records themselves began to lose the narrative structure that earlier characterized them.

Donating diagnostic blood from a vein, fingertip or ear lobe was a new experience for patients, but nothing captured the public's imagination quite like the X-ray, discovered exactly a century ago and used in medicine within weeks. Howell offers a thoughtful analysis of the assimilation of the X-ray into popular American culture as well as the professional implications of such questions as who pays for, takes and interprets the image. (One early specialist charged less



X-ray specs — fluoroscopy during a US Army operation in 1915. The picture's original caption ran "Intermittent control: surgeon and roentgenologist working simultaneously".

for normal than abnormal X-rays.) Curiously, despite the fact that both hospitals acquired their machines early, a few patients in the 1920s still presented with clinically diagnosed fractures and left hospital without an X-ray. By then, this was the rare exception, but patients and their doctors had to learn to adjust to the new technology, even so exciting an innovation as the X-ray. In the era before defensive medicine was being practised, it seemed unnecessary to some to use a machine to confirm a straightforward diagnosis such as a broken arm, and

simply having to take the patient to and from the equipment created new logistical problems within the hospital.

Howell is neither a technological apologist nor a wistful romantic. Technology changed the nature of the hospital during those decades, adding to its costs and making the surroundings less domestic and familiar. In many instances it changed neither the diagnosis nor the treatment the patient received, but sometimes it could, as when acute appendicitis was part of the differential diagnosis and when white blood cell counts mattered. Analysing a large random sample forces Howell to look at the ordinary, and it is at this level that his book is both intriguing and frustrating. Because his hospitals sometimes responded differently to various new technologies, he is acutely aware of the danger of generalizing from only two institutions. His original project aimed to compare the United Kingdom and the United States; the United

Kingdom is now reserved for a second volume and various US issues — about costs, or the development of paramedical groups, for example — are only cursorily examined. We are shown how the principles of accountancy began to influence hospital finance (the managerial revolution is of long duration), but the counter-intuitive revelation that employment costs apparently became less important between 1898 and 1920 (in one hospital, at least) passes without comment.

In his conclusion, Howell relates his historical research to contemporary ethical and economic debates. These lessons are not terribly surprising, but his early twentieth-century portrait is fresh and important, as an analysis of medical practice just coming to grips with the technological world. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

The shape of things to come

Ziauddin Sardar

Visions of the Future: The Distant Past, Yesterday, Today and Tomorrow. By Robert Heilbroner. *Oxford University Press*: 1995. Pp. 133. \$19.95.

The End of the Future: The Waning of the High-Tech World. By Jean Gimpel. *Adamantine*: 1995. Pp. 125. £32.50 (hbk); £14.95 (pbk).

THINKING about the future is a tricky and hazardous business. Tricky because our conventional way of thinking does not normally incorporate the future — we consciously have to strive to imagine what the future may hold, what anticipated and unexpected possibilities lurk on the distant horizon. Hazardous because the probability of getting one's forecasts wrong is very high.

It is not surprising then that both Robert Heilbroner and Jean Gimpel are rather cautious about making any predictions. Heilbroner is concerned chiefly with the way we think about the future and how this thinking has itself changed. Gimpel, on the other hand, focuses largely on the failures, unfulfilled promises and our misplaced optimism that technology would usher in a more humane and sane future. Both are economical with their analysis but nevertheless provide a punchy and thought-provoking examination of our inability to comprehend the future.

Heilbroner takes us through the history of humanity, as seen from the perspective of the West, at an exhilarating pace. The distant past, covering 150,000 years to yesterday, is marked by a wide range of diversity in all spheres of human existence, with

one exception: its view of the future. Until about 250 years ago, when yesterday began, the future was static; the present simply continued endlessly as there were no self-generating changes. The future alters with the arrival of capitalism, Western science and technology and popular political movements. Suddenly, endless progress becomes the main theme of the future. While the distant past continued for most of humanity, a new era appeared in the nations of the West. But faith in perpetual progress is now waning. Our trust in science and technology has been shaken. We are more sceptical about our political institutions and their ability to deliver a just and equitable society. Capitalism, we have learned, says Heilbroner, is only good for the rich. In the United States alone, the number of people living in poverty rose from 23 million to 35 million between 1975 and 1991, whereas the number of millionaires increased from 642 to 60,667. So tomorrow demands that we re-examine our basic assumptions.

Whereas Heilbroner asks that we rethink our economic principles, political institutions and the nature of our science and technology, Gimpel argues for a new way of seeing technological development. The crisis in the West is largely due to our disenchantment with a host of technologies that were seen purely in utopian terms. Such notions as the 'electronic cottage', intelligent and perfect computers and space travel were seen as panaceas that would rescue humankind from its follies. Instead, they have turned out to be mirages. Indeed, we are going backwards into the future: concerned citizens are giving up their cars

for bicycles, we increasingly prefer cotton, silk and linen to man-made fibres, and concrete is being abandoned in favour of timber-framed housing.

All this means that the West has reached a technological plateau. From now on, technologies may seem to be better, faster and more promising, but in reality they would not improve our lives, or deliver the greatest material benefits to most of humanity, or make us more happy. Moreover, we can already discern a marked decline in technological innovation. The belief in the power of technology to rescue our future is therefore dangerously obsolete.

The future is thus waiting to explode. We can no longer be sure, Heilbroner asserts, that we will actually survive. The only way to do so is to think more concretely and imaginatively about the future. In particular, we must focus our thinking about the future on three principles. First, we must ensure that we achieve a secure terrestrial base for life. This means abandoning many cherished notions about materialism, progress and perpetual advances in science and technology. Second, we must "find ways of preserving the human community as a whole against its warlike proclivities". This requires abandoning naked capitalism for more humane economics and seeking more appropriate models of governance. Third, the distant future must take us towards a much deeper and more abiding respect for 'human nature' and this can be ensured only if we make total respect for human dignity a prime focus in our contemporary cultural and educational concerns.

Gimpel does not offer a prescription for the future. He simply asserts that we have reached the end of the golden age of Western science and technology. And the West, like all other civilizations, as Ibn Khaldun and others have predicted, must decline in order to rise again in some far distant future. The present phase of the cycle of rise and decline of civilizations favours the Asian civilizations. The near future, the twenty-first century, belongs to Asia in general and to China in particular. The centre of world trade has already moved to the Pacific Basin. The nations of the Far East are growing at a phenomenal annual rate of eight per cent — and there is no indication that this will slow down in the next decades. The growth of Asia means the end of 'white supremacy' and the return to a future of a thousand years ago. "China is at the beginning of a cycle that could last a millennium, while western civilization stands at the end of a cycle that is already 1,000 years old", says Gimpel. Who could argue with that? □

Ziauddin Sardar, a consulting editor of Futures and visiting professor of science and technology policy at Middlesex University, is at 1 Orchard Gate, London NW9 6HE, UK.

Science literally speaking

Walter Gratzer

The Faber Book of Science. Edited by John Carey. Faber: 1995. Pp. 560. £17.50.

TAKE a close look, if you will, at this passage from an article in the *New Yorker* by Julian Barnes, whom I had counted among my favourite writers: "... both the present Prime Minister... and the present Leader of the Opposition have music-hall blood in their veins. That they should end up heading their parties in the House of Commons seems Darwinian: positive proof of the inheritance of acquired characteristics." You have to allow that Barnes's grasp of the principles of the theory of evolution is shaky (and that of the editorial staff of the *New Yorker* no less so, alas). I was shocked; but then I began to wonder whether, skulking in some underground chasm in Chicago or Geneva, one might not find a particle physicist who would have read Barnes's article and not even noticed. And I confess I have encountered successful biologists who are strangers even to the second law of thermodynamics (which was, you may recall, the shibboleth that helped C. P. Snow seek out the barbarians in our midst). Scientists, in short, do not always have deep roots in their culture.

All the more reason to wonder at the magisterial sweep of this remarkable compilation by John Carey, professor of English at the University of Oxford. *The Faber Book of Science* is everything an anthology should be; it is entertaining, stimulating and occasionally startling. Carey's reading is prodigious, and even the more familiar of his choices are illuminated by commentaries that crackle with literary and indeed scientific insights.

In his introduction, Carey contrasts what he sees as two types of writing about science; he calls these the mind-stretching (or jaw-slackening, which informs you how many of the water molecules that passed through Socrates in his hemlock draught you have in your breakfast cup of coffee or how far your DNA will stretch), and the explanatory, to which he is mainly drawn. He is tolerant of the occasional smugness of the Medawars and Haldanes, who, being scientists, are, in Medawar's words, not merely clever, but "have something to be clever about". He even sympathizes with Lionel Trilling, a man of letters, stricken by the bitter reflection that exclusion from the corpus of scientific thought was "bound to be experienced as a wound to our intellectual self-esteem". But Carey's is a mission of reconciliation, even to the extent that he finds a kinship between science and theology — a discipline that might, he believes, "be regarded as a

science". Here, I rather think, most of us would part company with him. The difference is surely encapsulated by Locke, who defined religious faith as something accepted not on the basis of evidence but on the authority of another. The cast of mind that draws scientists to their profession is the antithesis of that which predisposes towards religion. Most scientists, I suspect, would hold that not only is there no God, but the toast always falls on the carpet buttered-side down.

So then to the matter: Carey has arranged his selections chronologically by subject material, starting with Leonardo da Vinci's notebooks and progressing through the seventeenth and

Carey's reading is prodigious, and even the more familiar of his choices are illuminated by commentaries that crackle with literary and indeed scientific insights.

eighteenth centuries to the Victorian era, represented by majestic passages from works of the scientific Titans of the time, as well as such an unexpected figure in this company as John Ruskin with a sprightly essay on the nature of rust. Here are not only the crystalline periods of Lyell, but also the lugubrious stanzas that resulted when Tennyson read *Principles of Geology*.

Carey has trawled far beyond the usual confines of science writing. An account of August Kekulé's two celebrated dreams, which revealed to him the essence of stereochemistry and the structure of benzene, is culled from the *Journal of Chemical Education*, and for transfixing contemporary narratives about the discovery of X-rays and their properties he has gone to such sources as *McClure's Magazine* and *The Electrical Engineer* of 1896. By way of commentary he tells us that Professor Czermak of Graz could not sleep after seeing a photograph of his cranium which revealed to him the death's head beneath the skin, that inspecting the internal organs was widely held to be an affront to privacy, that Dr Banaduc of Paris exhibited X-ray photographs of the human soul and that M. Gaudoin of Dijon successfully offered high-dose irradiation as a depilatory beauty treatment. And he ends the section with a marvellous passage from Thomas Mann's *The Magic Mountain*.

The naturalists are well represented, not forgetting the romantics such as Maeterlinck, whose anthropomorphic

rhapsody on the life of the bees reads so quaintly today. There is a lethal parody by Mark Twain of Alfred Russel Wallace's proposition that the Universe was designed expressly to accommodate man (the strong anthropic principle, as it is now called), Albert Einstein and Arthur Stanley Eddington on the new physics, and a prizewinning exposition of the general theory of relativity by an Irish schoolteacher. (So much for Eddington's assertion that none but he and its begetter could properly understand the general theory.) Another unexpected pleasure is the limpid elegance of Max Born's inaugural lecture at Edinburgh on the subject of quantum mechanics.

Joseph Wood Krutch, professor of English at Columbia University and naturalist, is one of many unjustly forgotten writers whom Carey has disinterred, and he has discovered natural history among the essays of George Orwell and a small gem by John Steinbeck, whose interest was stimulated by his marine-biologist friend Ed Ricketts (portrayed as Doc in the Californian novels). And so to our own time, with verse and with the prose of Peter Medawar, Isaac Asimov, Arthur C. Clarke and many others.

An anthology is a distillation of one man's tastes and interests and it would be absurd to cavil about omissions, but I was surprised to find no H. G. Wells (if nothing else, then possibly his memorable portrait of T. H. Huxley at the blackboard in Kensington). Perhaps, like Max Beerbohm, Carey regards Wells's prose as so much "cold porridge spilt on the pavements of Gower Street". Nor are there represented here such outstanding contemporary writers on the ways of science and scientists as Jeremy Bernstein, Stephen Hall and Gary Taubes. But Carey has given us more than generous measure and no scientific anthologist that I know of, except, on a smaller scale, Martin Gardner, has even run him close.

Carey hints in his introduction at a more serious purpose than to entertain: "If young people are to be wooed back into science", he suggests, "it will not be done by telling them that if they continue to spurn it, Britain will face economic decline (true as that may be). But if scientists demonstrate by their writing that Medawar's promises of pleasure and self-fulfilment are true, they will not lack recruits." Let us pray that he is right, but he also contends that the true antithesis of science is politics, for this "is constructed out of preferences, which it seeks to elevate, by the mere manipulation of words, to the status of truths". Well now, in Britain at least (though not only here), politics has become, by its vulgar demand for instant economic returns, the most insidious enemy of science. The well is being poisoned. Rutherford used loudly to proclaim his

pity for those impoverished souls who had no labs to go to for solace and amusement. Science may be paying the price for letting the word leak out that scientists (the good ones at least) do it only because they enjoy it. But if that pleasure fades, I recommend Carey's admirable book as the next best thing. □

Walter Gratzer is at the MRC Muscle and Cell Motility Unit, King's College, 26-29 Drury Lane, London WC2B 5RL, UK. He is the editor of *A Bedside Nature: Genius and Eccentricity in Science 1869-1953*, an anthology soon to be published by Macmillan Magazines.

Social medication

Richard Davenport-Hines

The Pill: A Biography of the Drug that Changed the World. By Bernard Asbell. Random: 1995. Pp. 411. \$25.

THE contraceptive pill is an exceptional drug. For the past 30 years it has been swallowed as a daily routine by more humans than probably any other prescribed medication in the history of the world. Yet it is intended neither to prevent nor to cure an illness. The first medicine to have been designed for a social rather than a therapeutic purpose, it inaugurated what Bernard Asbell calls the Era of BioIntervention, in which the human reproductive system is regulated and re-devised in an unprecedented manner. For these and other reasons, the oral contraceptive has caused the most serious confrontation between the Roman Catholic Church and science since Galileo. Asbell's account of the scientific research preceding the marketing of an oral contraceptive, and its social sequel, is a balanced, readable and interesting mix of biography, scientific history, theology and public-policy analysis.

The most vivid biographical passages describe the lifetime's work of Margaret Sanger (1879-1966) and Katherine McCormick (1875-1967), whom he rightly identifies as "the indisputable mothers of the Pill". Sanger came from an Irish Roman Catholic family, and, as a girl at her mother's funeral, had furiously reproached her father: "You caused this. Mother is dead from having so many children." Afterwards, as a social worker in

New York's Lower East Side, she was revolted by "poor, weak, wasted, frail women, pregnant year after year like so many automatic breeding machines", and by the many deaths due to illegal abortions. She set out to provide women with reliable contraceptive information, and was imprisoned for her pains in 1916. Later she realized that she could raise large sums for her planned-parenthood projects and obtain protection from legal harassment if her rhetoric emphasized the dangers of over-population rather than focusing on sexual oppression.

One of her supporters was a beneficent multi-millionairess named Katherine McCormick, whose husband had been diagnosed as schizophrenic some years into their marriage and who had been convinced by Mendelian theory that she must never have children. Together, in 1951, Sanger and McCormick commissioned Gregory Pincus of the Worcester Foundation for Experimental Biology to develop an oral contraceptive that could be taken as easily as a pill of aspirin and that would have the same level of fallibility as vaccination. The motives of Sanger and McCormick were threefold: they wanted to simplify life for women, to distance sexual acts as much as possible from contraceptive acts and to reduce the



"Halfway to the corner... shawled, hatless... all day long and far into the evening in ever increasing numbers they came" — the first US birth-control clinic, opened by Margaret Sanger and her sister, Ethel Byrne, in 1916 at 46 Amboy Street, Brooklyn.

fertility of the 'unfit'.

Overall, McCormick provided \$2 million for the research. Pincus in 1951 proved that progesterone inhibited ovulation and began a new search for synthetic drugs. Asbell gives a lucid account of the work of other contemporary researchers in this area: the Harvard gynaecologist John Rock; the maverick chemist Russell Marker; the steroids researcher Carl Djerassi; and Frank B. Colton of the G. D. Searle pharmaceutical company. The upshot of all their efforts was the

marketing from 1957 by Searle of a synthetic anovulent, ostensibly to stop ovulation in women with menstrual disorders, although its contraceptive effects were so well known that by 1959 half a million women in the United States alone were using it.

In 1960 the US Food and Drug Administration accepted the Searle anovulent as an oral contraceptive pill — an event that, as the *Ladies' Home Journal* later declared, made more "immediate difference in women's lives" even than obtaining political suffrage. The universal attraction of this new product, and its importance to married and unmarried women, is indicated by the simplicity and ubiquity of its popular name in the English-speaking world, "the Pill" of Asbell's title. The Pill's appeal was not limited to women. Men were pleased to be relieved of contraceptive responsibility. Biochemical and hormonal contraception seemed like hard science to doctors and complemented their view of the necessity of birth-management by scientific experts.

From the beginning the Pill was detested by those who did not want the fear of pregnancy to be disconnected from the pursuit of sexual pleasure. Asbell gives a shrewd and temperate account of the background to the papal encyclical of 1968, *Humanae Vitae* ("Of Human Life"), which interdicted artificial birth control, as much as anything, one feels, from Pope Paul VI's fear that any change in teaching would seem to repudiate his papal predecessors. Asbell is rightly dismissive of the self-victimization of some American feminists who attacked oral contraceptives as devices that reduced them to passive objects of male lust.

The final section of *The Pill* will make ominous reading for all those who value the free traffic of scientific ideas. Asbell shows how contraception in the United States has become so politicized that its techniques have plummeted as subjects of research. Even the continuing research in the United States into pharmaceutical substances that will kill or disable sperm is conducted in an intimidating social and political context. The bigotry directed by religious fanatics against scientists and physicians is mediaeval in its primitive savagery. Thus the campaign by American zealots against the pharmaceutical company Hoechst, which has marketed the RU-486 abortive tablet since 1988, involved not merely the threat of a worldwide boycott of Hoechst products if RU-486 was sold outside France, but also chilling menaces of extortion and violence. It is dismaying that the leading Western democracy should increasingly be held hostage by ignorant and unscrupulous enemies of scientific freedom. □

Richard Davenport-Hines is at 51 Elsham Road, London W14 8HD, UK.

Profitable frontiers

Daniel S. Greenberg

The Golden Helix: Inside Biotech Ventures. By Arthur Kornberg. *University Science Books*: 1995. Pp. 287. \$28, £24.95. Distributed in Europe by W. H. Freeman.

THE Nobel laureate Arthur Kornberg does not hoard his spears in this candid, acerbic memoir of an academic scientist's venture into industry-financed research. Ravenous Wall Street investors, dense



Budd Gray

Talking candidly: Arthur Kornberg, with Lewis Thomas (right).

corporate management, arrogant scientists, misleading pharmaceutical advertising, confusion-sowing lawyers and myopic university administrators are plump targets for the professor emeritus of biochemistry at Stanford University's School of Medicine. In a calling rife with antagonisms, the bane of many professorial memoirs is a bland aversion to causing offence. Not Kornberg's.

He reports, for example, that in 1980, when he and the legendary scientist-entrepreneur Alex Zaffaroni, of Syntex renown but off on his own, sought capital for a start-up venture, the Syntex executives they visited "were utterly rude in their treatment of Alex, who had conceived and created the Syntex they occupied in Palo Alto". Kornberg criticizes Roy Vagelos, then president of Merck, for "disdain" for small biotech ventures, adding that "this disdain would dictate Merck policy for the next decade".

Johnson & Johnson's top-selling "Extra-Strength Tylenol", he caustically observes, "is merely a larger pill containing 500 grams of acetaminophen instead of the customary 375". Discussing Schering-Plough's frictional relationship in the late 1970s and early 1980s with its biotech-dependant, Biogen, Kornberg observes that the "scientists regarded the business people as shallow mercenaries, and the business people looked upon the scientists as a necessary evil". The difficulties, he writes, were heightened by "the

abrasive manner" in which board meetings were chaired by Walter Gilbert, a 1980 Nobel prizewinner. Gilbert, he reports, made "daily phone calls" to the chief executive officer, Robert Luciano, "berating the ineptitude of the Schering-Plough staff and the inadequacy of the company's operations".

Kornberg finds no end of targets. Scientists at the booming Genentech start-up company were "distracted by concerns about how best to enjoy and invest the bonanza profits pouring in from their ownership" of stock. And from his perspective as a witness in litigation challenging patent rights for the polymerase chain reaction (PCR) technique, Kornberg deplors "inappropriate commercial exploitation of knowledge that was paid for [by], and belonged to, the public".

Although corporate and academic stumbling and inanity are his backdrop, Kornberg is mainly concerned with telling the tale of a successful marriage of fundamental research and industrial money. Blessed with foresight and good luck, he says, he and his colleagues successfully

melded corporate support and university-style, open basic research in what he proudly describes and lauds as an exceptional creation: the DNAX Research Institute of Molecular and Cellular Biology, Inc., which, after a parched period, became a well-financed subsidiary of Schering-Plough. Focused on the molecular biology of T cells and cytokines, DNAX, founded in 1980, is described by Kornberg as "a small institute, remote from the imperatives of product development and marketing and given the freedom to innovate".

In these admirable respects, he asserts, it is a rarity in pharmaceutical research, as well as a miraculous survivor, given the ravenous demand for products and cash flow in the profit-starved biotechnology industry. Kornberg salutes the scientific and marketplace achievements of other biotech companies, notably Amgen and Genentech. DNAX has yet to produce an economic blockbuster, he acknowledges, but he expresses confidence that therapeutic value and riches are sure to ensue from Schering-Plough's development of DNAX's pioneering research on the powerful interleukin IL-10.

Steeped in the values and culture of academic fundamental research, Kornberg urges recognition of the continuing intellectual debt that the pharmaceutical industry owes to university-based science. "The industrial sector", he states, "must be reminded that the concepts, techniques, and practitioners of biotech

ventures and biotechnology driven pharmaceutical firms all came from basic academic research. As technologies proliferate and gain in sophistication, their scientific foundations are submerged and forgotten." And he warns against "illusions that basic research can be left to industry". "In fact", he declares, "more than 90 percent of such research has been, in the past, and must be in the future, done in university and other academic settings, requiring massive support to the tune of billions of dollars from the taxpayer through the federal government."

Kornberg's unexceptionable views on the wellsprings of scientific knowledge are embedded in a smartly written work that will reward readers in academe and industry. One puzzling dissonance, however, must be noted by this otherwise admiring reviewer. Yes, of course, as Kornberg writes, the indispensable billions for basic research in the United States must be provided by the taxpayers through the federal government. It is jarring, then, to find that elsewhere in *The Golden Helix*, Kornberg reports that Zaffaroni "incorporated DNAX in the Isle of Jersey, one of the Channel Islands, a maneuver designed to save taxes on future earnings as had been achieved by making Syntex a Panamanian company". No comment. □

Daniel S. Greenberg is editor and publisher of *Science & Government Report*, 3736 Kanawha Street NW, Washington DC 20015, USA.

Betrayer of truth

Anthony Snodgrass

Schliemann of Troy: Treasure and Deceit. By David A. Traill. *John Murray*: 1995. Pp. 365. £19.99.

DAVID Traill would hardly claim to be a detached observer of the excavator of Troy and Mycenae, his quarry for the past 15 years. This is the third book he has published, jointly or alone, on Heinrich Schliemann: with his painstaking search through diaries, newspapers, catalogues and weather-reports, he would make a great prosecuting counsel. He has proved beyond argument that Schliemann, if not a complete psychopath, was an unsavoury character, egocentric, a bully to the weak and a sycophant to the powerful, and, above all, profoundly dishonest.

Schliemann was not just economical with the truth: he purveyed lovingly crafted, circumstantial falsehoods especially but (alas) not only about his personal life. He was sometimes caught out in his lifetime, which explains the contempt in which some (not all) of his learned contemporaries held him, and

their unavailing, because misdirected, attempts to discredit his findings. The question today, a century and a quarter after his great discoveries began, is: how much does this still matter? Traill's case is that it matters very much. Once we realize that any statement of Schliemann's about the find-circumstances, location, manner and date of discovery of any portable object that he published (or even noted in his diaries) may be either a deliberate lie or a careless error, then the assemblages or (as he called them) treasures which did most to make him famous begin to fall apart.

In 1984, Traill showed that this was beyond doubt true of one of his most famous discoveries, the Early Bronze Age "Treasure of Priam", allegedly found on 31 May 1873 at Troy, presumably in one of the three or more different locations that Schliemann at different times attributed to it. Schliemann was careless enough to have illustrated some of the objects, later included in "Priam's Treasure", in the finds of earlier seasons. We also know that throughout his campaigns he was intermittently buying objects from local inhabitants. There is a suspicion that some of these may also have found their way into his assemblages.

In two chapters, Traill advances parallel arguments about the even more remarkable discoveries of November 1876, when Schliemann was excavating the Shaft Graves at Mycenae. The contents of these tombs have stood, ever since, as a landmark at the beginning of the Mycenaean civilization, belonging not (as Schliemann thought) to the time of the Trojan War, but three or four hundred years earlier, in the sixteenth and even

the seventeenth centuries BC. The tombs themselves are still there, and there is a reliable witness to the discovery of at least some of the most spectacular finds within them. But the rest? This time the evidence is almost all circumstantial: Schliemann did not repeat the mistake he made at Troy. But his account can be faulted for carelessness in transferring finds from one grave to another and, for the ten days that covered the excavation of the prodigiously rich Shaft Graves III, IV and V, maybe something worse.

The Shaft Graves span only a couple of generations, yet the diversity of style in the finds, even within the same tomb, might suggest a longer span. In at least one case (that of two terracotta figurines "from Shaft Grave I"), later knowledge has shown a definite discrepancy in date, of at least 200 years. Traill argues that, for weeks before the finding of the Shaft Graves, Schliemann had been excavating burials, some of them directly above (and therefore, like the figurines, later than) the Shaft Graves themselves.

Worse still, if there is one find that has become emblematic for Schliemann, and for the prehistoric Aegean generally, it is the moustachioed and bearded gold mask, still often and mistakenly dubbed the "Mask of Agamemnon", which duly appears on the jacket of this book. With its up-turned moustache and imperial goatee, it is utterly unlike the other four gold masks found in the same grave and its neighbour — why? Traill states that examination of the mask shows the moustache to have been originally shown as drooping: at some point it was "rather crudely" altered. Was this in the sixteenth century BC or in the 1870s, between the reigns of Napoleon III and those of Tsar Alexander III and Kaiser Wilhelm II (to name just three rulers that it vaguely recalls)? One would like more confirmation — say, a death-



From Schliemann of Troy

Lost and found: "Priam's Treasure", attributed to 1873. Several of the pieces were in fact found in 1871 or 1872.

bed confession from the hypothetical goldsmith responsible for this and other work, such as the mass-reproduction of some of the 701 gold discs found only in Grave III. But it is disquieting enough to make one ponder the reality, scale and homogeneity of the "Shaft Grave culture" that has dominated every account of the period since.

Yet Schliemann's discoveries, however much embroidered and added to by his dishonesty, were real enough: the buildings at least, even older than he himself thought them, are there for all to see. Even when we can catch him red-handed 'salting' his discoveries, the intrusive objects mostly turn out to belong to the right period (a point that Traill does not fully face). In the end, for all the damaging points that Traill makes, the abiding source of worry is a different one from his. Schliemann conformed to a type that one recognizes at once from the fringes of any science, or from among its weaker students — ingratiating, exaggeratedly deferential to scholarly reputation, ready to borrow without acknowledgement from other books in his own writing: someone whom, if he laid claim to a momentous discovery, one would instantly assume to have fabricated it. This is exactly how many of his contemporaries thought of Schliemann, and, in the last analysis, they were wrong. □

Anthony Snodgrass is at the Faculty of Classics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK.



From Schliemann of Troy

Telling it as it isn't: Schliemann speaking to the Society of Antiquaries at Burlington House, 1877.